Available online at www.sciencedirect.com
ScienceDirect
Renewable and Sustainable Energy Reviews
11 (2007) 551–602
RENEWABLE & SUSTAINABLE ENERGY REVIEWS
www.elsevier.com/locate/rser

Discourses on solar radiation modeling

T. Muneer*, S. Younes, S. Munawwar

School of Engineering, Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, UK

Received 13 April 2005; accepted 5 May 2005

Abstract

All solar energy applications require readily available, site-oriented and long-term solar radiation data. A typical database comprises of global, direct and diffuse solar irradiance, duration of sunshine and complementary data like cloud cover, atmospheric turbidity, humidity, temperature, etc. However, most of these stations do not provide complete if any information on solar data, chiefly due to the capital and maintenance costs that measuring instruments incur. For instance, global radiation is the most frequently measured parameter, its two components, i.e. diffuse and direct irradiance are often not measured.

Improvements have been made to the meteorological radiation model MRM which, had been developed by Muneer et al. as a simple broadband irradiance estimation model based on synoptic information, by incorporating the sunshine information in the model's regressions. The result of the improvement of the model is a considerable reduction in biases and scatter in the comparison between estimated and measured data. The improved meteorological radiation model, IMRM is more accurate, by up to 70% in some cases, than its predecessor in estimating, global and diffuse horizontal irradiance.

When sunshine, atmospheric pressure and temperature are not measured by a nearby station, yet cloud information is recorded, radiation estimation models based on cloud cover, CRM, can be used. Three CRMs have been compared to newly proposed models. It was found that models with locally fitted coefficients gave a more accurate estimation of the solar radiation than CRMs with generalized coefficients. The newly proposed model performed better that the older generation models.

The third section of the article deals with estimation of diffuse radiation and possible improvements in its modeling. In this section, apart from clearness index (k_t), influence of the synoptic parameters of sunshine fraction (SF), cloud cover (CC) and air mass (m) on diffuse fraction of global radiation (k) is studied both qualitatively and quantitatively. It is found that, SF shows a strong bearing on the k–k_t relationship followed by CC and then m. As a next step, a series of models are developed for k as a polynomial function of k_t, SF, CC and m. After an extensive evaluation

*Corresponding author. Tel.: +44 131 455 2541; fax: +44 131 455 2264.
E-mail address: t.muneer@napier.ac.uk (T. Muneer).

doi:10.1016/j.rser.2005.05.006

Nomenclature

I_E	extraterrestrial horizontal irradiance
I_G	global horizontal irradiance
I_{GC}	clear-sky global horizontal irradiance
I_D	diffuse horizontal irradiance
I_B	beam horizontal irradiance
k_t	clearness index
k	diffuse to global ratio
DBR	diffuse to beam ratio
K_b	beam clearness index
$I_{G,C}$	Page clear-sky global horizontal irradiance
$I_{D,C}$	Page clear-day diffuse horizontal irradiance
$I_{D,OC}$	Page overcast-day diffuse horizontal irradiance
SF	sunshine fraction
P_{atm}	atmospheric pressure
DBT	dry bulb temperature
WBT	wet bulb temperature
RH	relative humidity
N	cloud cover (octa)
m	air mass
τ_r	Rayleigh scattering transmittance
τ_α	Mie scattering transmittance
τ_g	mixed gases scattering transmittance
τ_o	ozone scattering transmittance
τ_w	water vapor scattering transmittance
COF(i)	coefficients used in MRM and IMRM
HDD	hourly to daily datasets
HSHD	hourly or sub-hourly datasets
σ	standard deviation
$\bar{k}$	weighted mean of k
α	solar altitude (radian)
MBE	mean bias error (W/m^2)
RMSE	root mean square error (W/m^2)
R^2	coefficient of determination
Slope	slope of best fit line
abs	absolute value
I'	estimated value of the given irradiance component
I	measured value of the given irradiance component
I_m	mean value of the given irradiance component

procedure, a regression model is selected such that the diffuse radiation can be estimated with reasonable accuracy without making the model overtly complex. It was found that among all the models, the composite model involving all parameters provides the most accurate estimation of diffuse radiation. The site-specific models are further investigated for any appreciable correlations between different locations and their possible attributions. It was also found that a single model

could more than adequately estimate the diffuse radiation for the locations within a given region. Three optimum models are also recommended for each region, in view of the fact that information on all parameters is not necessarily available for all sites. This study reveals a significant improvement from the conventional k–k_t regression models to the presently proposed models, therefore, leading to more accurate estimation of diffuse radiation by approximately 50%.

Keywords: Solar horizontal irradiance; Quality control; Modeling

Contents

1. Introduction 554
2. Data assimilation 556
3. Quality control of solar radiation datasets 558
   3.1. The need for quality control 558
   3.2. The quality control procedure 560
4. Statistics for model comparisons 561
   4.1. Statistical indicators 561
   4.2. Accuracy score 562
5. Meteorological radiation model 563
   5.1. A brief review of the all-sky meteorological radiation model 563
   5.2. Improvements on MRM (IMRM) 564
   5.3. Validation of the IMRM 565
   5.4. Discussion of results 567
6. Cloud radiation model 568
   6.1. Review of available cloud cover based radiation estimation models 568
   6.2. Analyzing cloud cover distribution 572
   6.3. Improved multivariate models and hybridization 574
   6.4. Discussion of results 575
7. Diffuse radiation model 582
   7.1. Potential for improvement in the k–k_t regression model for the estimation of diffuse radiation 582
   7.2. Bearing of sunshine fraction, cloud cover and air mass on estimation of diffuse radiation 586
      7.2.1. Regressions under sunshine fraction (SF) bands 587
      7.2.2. Regressions under cloud cover (CC) bands 588
      7.2.3. Regressions under air mass (m) bands 589
   7.3. Modeling of diffuse radiation: the present approach 589
      7.3.1. Effect of clearness index on diffuse ratio 589
      7.3.2. Effect of sunshine fraction, cloud cover and air mass on diffuse ratio: individual basis with clearness index as the common parameter 590
      7.3.3. Effect of clearness index, sunshine fraction, cloud cover and air mass on diffuse ratio: combined basis 590
   7.4. Evaluation of developed models 591
      7.4.1. D_m vs. D_c plots 591
      7.4.2. Error histograms 591
   7.5. Region-wise recommendation of optimum models 593
8. Conclusions 599
References 601

All irradiance values are in W/m^2, all ratios and indexes are dimensionless.

1. Introduction

Knowledge of solar radiation data is a pre-requisite for the simulation and design of all solar energy systems. Architects, agriculturalists, air conditioning engineers and energy conscious designers of buildings also require such information. Most of the solar energy applications involve tilted surfaces. In order to effect the estimation of radiation on tilted surfaces, knowledge of both diffuse and direct components of global radiation falling on a horizontal surface is required.

However, quite often, projects involving utilization of solar energy are not supported by the required solar data at the place of interest, mainly due to the capital and maintenance costs that measuring instruments incur. It was observed that solar radiation is not as commonly measured as the respective elements across the globe. For instance, in the UK, some 600 sites measure synoptic information such as temperature, relative humidity, atmospheric pressure along with cloud cover, out of which 230 sites measure sunshine duration while solar radiation is measured at relatively fewer sites.

There are a number of problems that arise in the measurement of the synoptic and solar radiation. The most common causes of the errors are related to the equipment and their respective sensitivities. Another major cause of errors is site operation problems, such as instrument proximity to shading elements, electrical and magnetic fields, bird and insect activity and weather elements. There is, therefore, the need to identify erroneous data and exclude or correct them. A quality control procedure was developed, based on physical and statistical tests to remove suspected outliers in the datasets. The available datasets were used to fine tune the proposed models.

When solar radiation data is unavailable, it is possible to get reasonable accurate radiation estimates using the proposed models which rely on alternative synoptic information. These models are also useful to fill-in any gaps in the measured radiation datasets.

An improved meteorological radiation model IMRM has been developed as a simple broadband irradiance estimation model based on synoptic information. IMRM utilizes four measured synoptic variables namely, atmospheric pressure, dry bulb temperature, sunshine fraction/ duration and wet bulb temperature. The model is based on regressions between the ratio of diffuse irradiance to beam irradiance, DBR, and beam clearness index, k_b. IMRM calculates the beam horizontal irradiance, then the diffuse horizontal irradiance. Global horizontal irradiance is then deduced from its two components. The accuracy improvement of IMRM compared to the original MRMwas quantitatively evaluated using statistical indicators.

When only cloud information is recorded, radiation estimation models based on cloud cover, CRM, can be used. Three old generation CRMs have been compared to newly proposed models. The older models are based on regressions between the solar horizontal irradiance, solar geometry and cloud cover. Improvements in the regressions were made to two of the newer models, while an additional variable, air mass, was added in the last of the newer model. The new CRMs calculate I_G and I_B independently, and then I_D is deduced based on the previously calculated components. The different models were compared using accuracy indicators to identify the best performing model.

Acknowledgement of the fact that diffuse radiation is an integral component in calculating solar radiation falling on a tilted surface and the infrequency with which it is being measured, led to the evolution of a number of regression models for their estimation. Over the past, several regressions have been proposed between the diffuse fraction of global radiation and clearness index (ratio of global to extraterrestrial radiation) based on hourly, daily and monthly averaged data. Besides global radiation, it was found that other weather elements could also significantly effect and thus improve the diffuse irradiance estimation. Therefore, under the scope of present work, apart from k_t, daily SF (ratio of number of bright sunshine hours to day length) CC and m are both qualitatively and quantitatively investigated as to explore their significance in improved modeling of diffuse irradiance. The first part of the study investigates their influence, if any, on the performance of diffuse radiation models. In other words, each parameter is analyzed individually in view of its bearing on k–k_t relationship through graphical means. The second part of the study involves the development of models for k (diffuse ratio) based on k_t, SF, CC and m. For any given grouping of the above parameters, linear, quadratic and cubic models are tested. A detailed analysis of all the models is carried out through statistical means. Diffuse radiation is calculated from the developed models and compared against the measured values. An optimum model is selected which yields reasonable accuracy without making the estimation procedure overtly complex.

As an aid to understand when to use each of the above mentioned models, Fig. 1 associates a solar radiation model to each of the possible classifications of terrestrial meteorological measurement stations. The stations range from the simplest being a remote

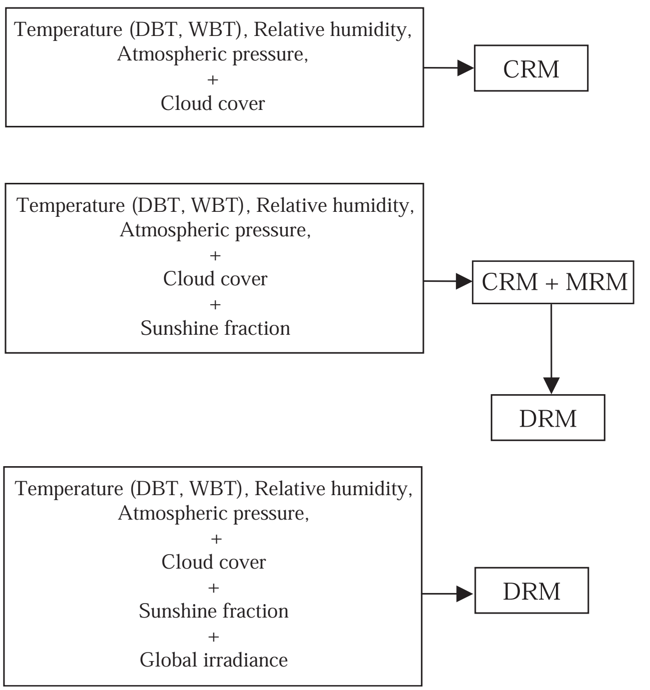


Fig. 1. Availability of alternate meteorological datasets and their use in generating solar radiation.

installation measuring temperature, relative humidity, atmospheric pressure, wind speed and direction and rainfall to the most complex first class measuring stations, where in addition to the former mentioned components, sunshine duration, cloud cover and various aspects of solar radiation are measured.

2. Data assimilation

The datasets available for the present analysis cover three countries, each with specific meteorological and geographical characteristics. Hubbard [1] demonstrated that the length of data series should be more than one year to characterize the seasonal pattern in special variability. Gueymard [2] recommends that in order to validate radiation estimation models, a minimum of three years datasets are needed. With the exception of Bahrain, all the datasets presently used span over a period of three or more years.

Two datasets were taken from UK locations, Aldergrove (54.65N; 6.25W) and Bracknell (51.26N; 0.45E), with a 5-year (1990–1994), and 6-year (1990–1995) period, respectively. The two UK locations are an example of temperate, maritime weather found in northern Europe. The two, radiation and meteorological measurement, stations are classed as 1st class stations, and are part of the UK Meteorological Office network. The radiation is recorded hourly from minute-by-minute averaged data. The cloud cover is recorded hourly.

Data from two Southern European sites were utilized, both sites being from Spain, i.e. Gerona (41.97N; 2.88E) and Madrid (40.45N; 3.73W). The Gerona dataset spanned seven years, covering the period from 1995 to 2001 and was provided by the University of Gerona, while the Madrid dataset spanned three years, covering 1999–2001 years. For both Spanish sites cloud cover data are recorded at 07:00, 13:00 and 18:00 h local clock time. The Instituto National De Meteorologia based in Madrid provided these data.

Data from three Indian sites were also utilized in the present work. All of these datasets were provided by the Indian Meteorological Department based in Pune. Data from Indian sites—Chennai (13.0N; 80.18E), Mumbai (19.12N; 72.85E) and Pune (18.53N; 73.85E) spanned 5 years covering 1990–1994 and Delhi (28.60N, 77.20E) spanning 1989–1998. As with the Spanish sites, the Indian cloud cover data were also not recorded hourly. Rather this information was noted at 08:30, 11:30, 14:30 and 17:30 h local time.

A dataset from the middle-east, Bahrain (26.22N; 50.65E) was used. This datasets provides hourly solar radiation data and sunshine duration from 2000 to 2002. And finally two Japanese sites, Fukuoka (33.52N; 130.48E) and Sapporo (43.05N; 141. 33E) were used. Each dataset spanned 3 years, 1992–1994 for Fukuoka and 1991–1993 for Sapporo. These two datasets are Courtesy of Kyushu University, Japan. They included global horizontal irradiance and both diffuse and beam components and the sunshine fraction on 10 min intervals.

In Section 5, the datasets were split into two categories. The Hourly to sub-hourly datasets (HSHD) are for locations that provide the same measurement frequency for sunshine fraction and the corresponding irradiance data. Hourly to daily datasets (HDD) are for locations that provide hourly irradiance data corresponding to the daily sunshine fraction data.

In Section 6, since the solar radiation models are based on cloud cover, Aldergrove, Bracknell, Chennai, Gerona, Madrid and Pune are the only seven datasets that were used since they included the cloud cover.

Table 1
Table shows the synoptic and radiation information of the datasets used in this study

Location	Radiation Measured (G = global, D = diffuse, B = beam)	Temperature (DBT, WBT)	Radiation measurement Frequency (Hourly/ 10 minute)	SF measurement Frequency (Daily/Hourly/ 10 minute)	Cloud Cover Frequency (Daily/Hourly/ Spot values)
Aldergrove	G,D	none	Hourly	Hourly	H
Bracknell	G,D,B	D,W	Hourly	Hourly	H
Chennai	G,D	none	Hourly	Daily	S: 0830, 1130, 1430, 1730
Fukuoka	G,D,B	D,W	10 min	10 min	None
Gerona	G,D	none	Hourly	Daily	S: 0700, 1300, 1800
Madrid	G,D	none	Hourly	Daily	S: 0700, 1300, 1800
Manama	G,D	D,W	Hourly	Hourly	none
Mumbai	G,D	none	Hourly	Daily	S: 0830, 1130, 1430, 1730
New Delhi	G,D	none	Hourly	Daily	S: 0830, 1130, 1430, 1730
Pune	G,D	none	Hourly	Daily	S: 0830, 1130, 1430, 1730
Sapporo	G,D,B	incomplete	10 min	10 min	none

Table 2
Table shows the geographical information, the length and classification of the datasets used

Country	Location	Latitude; longitude	Year	Description of dataset
Bahrain	Manama	26.22N; 50.65E	2000–2002	HSHD
India	Chennai	13.00N; 80.18E	1990–1994	HDD
India	Mumbai	19.12N; 72.85E	1990–1994	HDD
India	New Delhi	28.60N; 77.20E	1988–1998	HDD
India	Pune	18.53N; 73.85E	1990–1994	HDD
Japan	Fukuoka	33.52N; 130.48E	1992–1994	HSHD
Japan	Sapporo	43.05N; 141.33E	1991, 1993	HSHD
Spain	Gerona	41.97N; 2.88E	1995–2001	HDD
Spain	Madrid	40.45N; 3.73W	1999–2001	HDD
UK	Aldergrove	54.65N; 6.25W	1990–1994	HSHD
UK	Bracknell	51.26N; 0.46W	1990–1995	HSHD

HSHD refers to hourly or sub hourly radiation and sunshine duration dataset. HDD refers to hourly radiation to daily sunshine duration dataset.

In Section 7, diffuse radiation modeling was demonstrated for nine of the above sites with the exception of Aldergrove and Bahrain.

Tables 1 and 2 provide details on the content of the datasets.

3. Quality control of solar radiation datasets

3.1. The need for quality control

Any likely sources of errors or problems related to solar radiation measurement may be categorized under the following two major categories: equipment error and uncertainty and operation related problems and errors.

With any measurement there exist errors, some of which are systematic and others inherent of the equipment employed. The most common sources of error arise from the sensors and their construction. These are broken down into the most general types of errors and described below:

- Cosine response
- Azimuth response
- Temperature response
- Spectral selectivity
- Stability
- Non-linearity
- Shade-ring misalignment
- Dark offset (nocturnal) long-wave radiation error

Of all the above listed errors the cosine effect is the most apparent and widely recognized.

This is the sensor's response to the angle at which radiation strikes the sensing area. The more acute the angle of the sun, i.e. at sunrise and sunset, the greater the error will be (at altitude angles of sun below 6°). Cosine error is typically dealt with through the exclusion of the recorded data at sunrise and sunset times.

The azimuth error is a result of imperfections of the glass domes and in the case of solarimeters the angular reflection properties of the black paint. This is an inherent manufacturing error, which yields a similar percentage error as the cosine effect.

Like the azimuth error the temperature response of the sensor is an individual fault for each cell. The photometers are thermostatically controlled hence the percentage error due to fluctuations in the sensor's temperature is reduced. The pyranometers rely on their construct, i.e. a double-glass envelope to prevent large temperature swings.

The spectral selectivity of the pyranometers is dependent on the spectral absorbance of the black paint and the spectral transmission of the glass. The overall effect contributes only a small percentage error to the measurements. Each sensor possesses a high level of stability with the deterioration of the cells resulting in approximately $\pm 1\%$ change in the full-scale measurement per year. Finally the non-linearity of the sensors is a concern especially with photometers. It is a function of illuminance or irradiance levels. It however tends to contribute only a small percentage error towards the measured values.

The work undertaken by NREL [3] under the US continental climate and a desert site in Saudi Arabia has shown that zero offsets of -5 to $-20\,\mathrm{W/m^2}$ occur in diffuse pyranometer measurements due to thermal imbalances. This error was reported for all instruments using black sensors. The offset for a black and white detector, however, was found to be insensitive to such offset errors.

In addition to the above sources of equipment-related errors, care must be taken to avoid operational errors highlighted below:

- Operation related problems and errors
- Complete or partial shade-ring misalignment
- Dust, snow, dew, water-droplets, bird droppings, etc.
- Incorrect sensor leveling
- Shading caused by building structures
- Electric fields in the vicinity of cables
- Mechanical loading of cables
- Orientation and/or improper screening of the vertical sensors from ground-reflected radiation
- Station shut-down

The sources of operation relation errors itemized above are self-explanatory. It is good practice to protect cables from strong electric fields such as elevator shafts. Another source of error that may arise is from cables under mechanical load (piezoelectric effects). The piezoelectric effect is the production of electrical polarization in a material by the application of mechanical stress. Failure to protect cables from the above sources may produce 'spikes' in the data and these are shown as unusually high values of irradiance. Such errors are best highlighted via cross plotting diffuse ratio ($k = I_D/I_G$) against clearness index ($k_t = I_G/I_E$), and a sample plot is shown in Fig. 2. Note that any consistent errors emanating from an operational problem, such as misaligned shade ring are easily pickedup by this type of plot. Any serious departure of data from the normally expected envelope is thus identified.

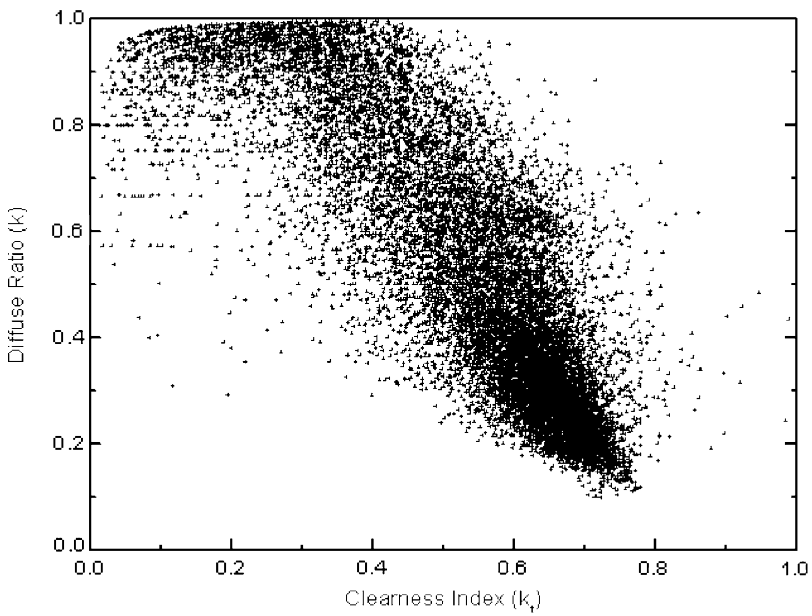


Fig. 2. Diffuse ratio—clearness index plot for Mumbai.

Stoffel et al. [3] give us a good representation of the scale of errors for carefully managed irradiance sensors. For their study, they found that the range of error for a pyranometer compared with a reference pyranometer was from $+2.5$ to -10%; while for a pyrheliometer the range was $\pm 2.5\%$.

3.2. The quality control procedure

The hourly irradiance data were quality controlled using the Younes et al. [4] algorithms. This procedure is based on a series of physical tests and a semi-automated statistical test. The physical tests are a series of three logical procedures and are as follows: First test: entries that show a solar altitude less than 7° are eliminated. This procedure aims to remove sunrise and sunset data that is affected with cosine error. Although researchers cannot agree on a particular value beyond which the radiation is data is free of the effects of cosine response, anywhere is the range of 6–10° is acceptable.

The second test is a logical test, as the clearness index and the diffuse ratio are always positive and have values between 0 and 1. Negative values and values that are greater than one are physical impossibilities, and therefore, imply measurement errors, therefore, $0 \leq I_{\mathrm{D}} \leq I_{\mathrm{G}} \leq I_{\mathrm{E}}$.

$$0 < k_t < 1 \text{ and } 0 < k < 1 \tag{1}$$

At this stage, the third test, the global and diffuse irradiances are compared with their corresponding Page-model upper and lower boundaries [5]. The global horizontal irradiance aught to be less than or equal to the clear-day global horizontal irradiance. Thus,

$$I\mathrm{G} \leq I_{\mathrm{G,C}} \tag{2}$$

From Muneer-Fairooz [6], it is proposed to test if the diffuse horizontal irradiance is sandwiched between the clear-day diffuse and the overcast-day horizontal irradiance as defined by Page [5]. Thus,

$$I_{\mathrm{D,OC}} \leq I_{\mathrm{D}} \leq I_{\mathrm{D,C}} \tag{3}$$

The statistical test is based on a standard deviation procedure in a k_{t}–k scatter plot. k_{t} is the clearness index ($= I_{\mathrm{G}}/I_{\mathrm{E}}$) and k is the diffuse to global irradiance ratio ($= I_{\mathrm{D}}/I_{\mathrm{G}}$). From the original work by the authors, it was found that each dataset had an optimum value of standard deviation to describe the quality controlled envelope of the data. For example, the Aldergrove envelope was found to have the $\bar{k} \pm 2.0\,\sigma_{\mathrm{k}}$ optimum quality control envelope of acceptance. The boundaries of the envelope are described mathematically. The upper and lower boundaries, are respectively, represented as,

$$A(k_{\mathrm{t}}) = \mathrm{Max}\left(1, a_1 k_{\mathrm{t}}^2 + b_1 k_{\mathrm{t}} + c_1\right) \tag{4}$$

$$B(k_{\mathrm{t}}) = \mathrm{Min}\left(0, a_2 k_{\mathrm{t}}^2 + b_2 k_{\mathrm{t}} + c_2\right) \tag{5}$$

A cut-off is needed in most cases in the lower k_{t}—upper k region, where the lower boundary B (k_{t}) would remove good data. This is shown as C (k_{t}). The mathematical envelope of acceptance can be observed in Fig. 3, for the Aldergrove dataset.

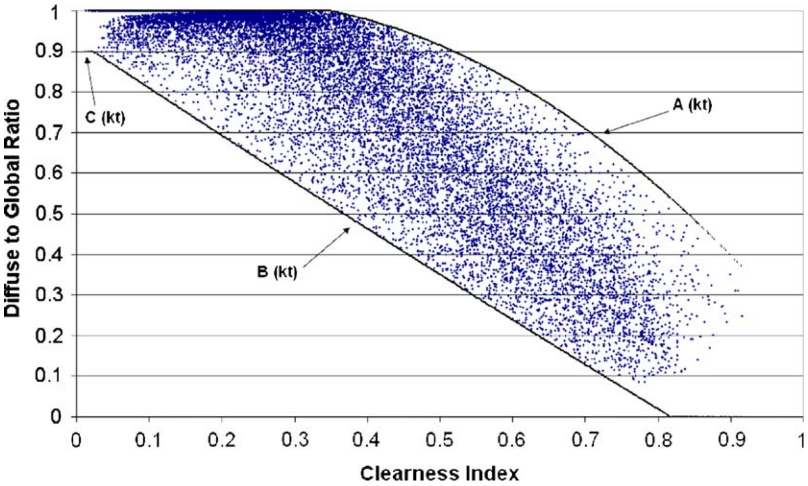


Fig. 3. Quality Control boundaries in a k_t–k plot for Aldergrove.

4. Statistics for model comparisons

4.1. Statistical indicators

The following statistical indicators were used in the following sections to quantitatively evaluate the performance of each of the models.

The slope of the best fit line between the computed and measured variable is desired to be equal to one. Slope values exceeding one indicate over-estimation, while slope values under one indicate under-estimation of the computed variable.

The coefficient of determination (R_2) is the ratio of explained variation to the total variation. It lies between zero and one. A high value of R_2, thus indicating a lower unexplained variation, is desirable. R2 is often used to judge the adequacy of a regression model, but it should not be the sole criterion for choosing a particular model as the value of R_2 increases with the number of coefficients in the model.

$$R^2 = \frac{\sum(I' - I_{\mathrm{m}})}{\sum(I - I_{\mathrm{m}})} \tag{6}$$

The MBE provides an indication of the trend of the model, whether it has a tendency to under-predict or over-predict its modeled values. MBE can be expressed either as a percentage or as an absolute value. Nevertheless, within a data set overestimation of one observation can cancel underestimation of another. An MBE nearest to zero is desired. It is given by the following equation:

$$\mathrm{MBE} = \frac{\sum(I' - I)}{n} \tag{7}$$

The RMSE gives a value of the level of scatter that the model produces. This is an important statistical test, as it highlights the readability and repeatability of the model. It provides a term-by-term comparison of the actual deviation between the predicted and the measured values. Since it is a measure of the absolute deviation, RMSE is always positive.

A lower absolute value of RMSE indicates a better model. Mathematically, it is given by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum (I' - I)^2}{n}} \tag{8}$$

Note that in Eqs. (6)–(8), I' is the estimated—I the measured and I_{m} the mean value of the given irradiance component (I_{G} or I_{D}) and n the number of data points.

Skewness is defined as a measure of the lack of symmetry in a distribution. A positively skewed distribution tails off to the high end of the scale while negative skew tails off the low end of the scale. If the distribution is normal or, in other words, has no skewness, then the skewness statistic will be zero. This will indicate a robust model.

Kurtosis is defined as a measure of the degree of peakedness in the distribution, relative to its width. The kurtosis statistic will be zero (mesokurtic) for a normal distribution, positive for peaked distributions (leptokurtic) and negative for flat distributions (platykurtic). A leptokurtic distribution of the errors is highly desirable. A high positive value of kurtosis represents that there are fewer outliers in the estimation.

4.2. Accuracy score

The Accuracy score procedure, referred as AS, aims to facilitate a discrete comparison between all the models. Numerically, it is expressed as:

$$\begin{aligned}\mathrm{AS} = {} & \frac{R_i^2}{R_{i,\max}^2} + \left[1 - \frac{|\mathrm{MBE}|_i}{|\mathrm{MBE}|_{i,\max}}\right] + \left[1 - \frac{|\mathrm{RMSE}|_i}{|\mathrm{RMSE}|_{i,\max}}\right] \\ & + \left[1 - \frac{|\mathrm{Skewness}|_i}{|\mathrm{Skewness}|_{i,\max}}\right] + \frac{|\mathrm{Kurtosis}|_i}{|\mathrm{Kurtosis}|_{i,\max}} + \left[1 - \frac{\Delta\mathrm{Slope}_i}{\Delta\mathrm{Slope}_{i,\max}}\right]\end{aligned} \tag{9}$$

$$\Delta\mathrm{Slope}_i = \left|1 - \mathrm{Slope}_i\right| \tag{10}$$

$$\Delta\mathrm{Slope}_{i,\max} = \left|1 - \mathrm{Slope}_{i,\max}\right| \tag{11}$$

Slope_i is the slope of the best fit line between the computed and measured irradiance component. The subscript i, max indicates the largest value of the given parameter for all of the models.

AS is a convenient index by means of which it is possible to compare the performance of any suite of models. Note that for the above mix of statistical parameters it is not necessary that a clear 'best' model may emerge owing to the fact that a gain in accuracy indicated by, say a low RMSE may be offset by a high value of kurtosis. The AS scoring system is thus of good value when such a contradictory picture emerges. The model that yields the highest values of R^2 and kurtosis would have the highest score. For MBE, RMSE and skewness, the model that yields the values closest to 0 would score the highest. Similarly, for the slope of the best fit line, the model that yields a score closest to 1 would score the highest. The maximum obtainable score per statistical indicator is 1, therefore, per solar irradiance component; a model would have a maximum obtainable score of 6. In Section 7, the diffuse radiation model uses a reduced version of AS, that does not take account of the slope as a statistical indicator. Therefore, the maximum obtainable AS per irradiance component is five. It should be noted that obtaining the maximum score does not

necessarily imply that the model is accurate and best performing; it only indicates that the model yields better results compared to the other models in the evaluation.

5. Meteorological radiation model

5.1. A brief review of the all-sky meteorological radiation model

The meteorological radiation model (MRM) has been developed by Muneer et al. [7–10] as a simple broadband irradiance estimation model based on synoptic information.

The MRM utilizes commonly measured variables namely, atmospheric pressure (P_{atm}), dry bulb temperature (DBT), sunshine fraction/duration (SF/SD) and wet bulb temperature (WBT). WBT can be obtained via DBT and relative humidity (RH), if direct measurements of the former element are not recorded.

The MRM is based on regressions between the ratio of hourly diffuse horizontal irradiance (I_D) to beam horizontal irradiance (I_B) and beam clearness index. Note that the above two quantities are herein referred as $DBR = I_D/I_B$ and $K_B = I_B/I_E$, where I_E is the extraterrestrial horizontal irradiance. Muneer et al. [8–10] have expressed the relationship between the above two-dimensionless variables in the form of a power function:

$$DBR = a(K_B)^b \tag{12}$$

They validated the MRM using data from the UK and Japan; the coefficients to use in Eq. (12) for UK are $a = 0.285$ and $b = -1.00648$.

The calculated beam horizontal irradiance (I_B) is a function of the extraterrestrial horizontal irradiance attenuated by the sunshine fraction (SF) and atmospheric transmittances thus,

$$I_B = I_E \mathrm{SF} \tau_r \tau_\alpha \tau_g \tau_o \tau_w \tag{13}$$

τ_r and τ_α are transmittances due to Rayleigh and Mie scattering; τ_g, τ_o and τ_w the transmittances due to mixed gases, ozone and water vapor scattering, respectively. These are expressed mathematically as,

$$\tau_\alpha = \exp\left[-k_\alpha^{\mathrm{COF}(1)}\left(1 + k_\alpha - k_\alpha^{\mathrm{COF}(2)}\right) m^{\mathrm{COF}(3)}\right] \tag{14}$$

$$\tau_o = 1 - \left[0.1611 x_o (1 + 139.48 x_o)^{-0.3035} - 0.002715 x_o \left(1 + 0.044 x_o + 0.0003 x_o^2\right)^{-1}\right];$$

$$x_o = l_o m \tag{15}$$

$$\tau_r = \mathrm{COF}(4) - \mathrm{COF}(5) m + \mathrm{COF}(6) m^2 - \mathrm{COF}(7) m^3 + \mathrm{COF}(8) m^4 \tag{16}$$

$$\tau_w = 1 - \mathrm{COF}(9) x_w \left[(1 + \mathrm{COF}(10) x_w)^{\mathrm{COF}(11)} + \mathrm{COF}(12) x_w\right]^{-1}; \quad x_w = l_w m \tag{17}$$

$$\tau_g = \exp\left(-\mathrm{COF}(13) m^{\mathrm{COF}(14)}\right) \tag{18}$$

Refer to Muneer et al. [7–10] for the values of above used coefficients for UK sites.

Once I_B is calculated, I_D is then calculated via Eq. (12). The calculated beam and diffuse horizontal irradiance are then summed up to obtain the calculated global horizontal irradiance (I_G).

5.2. Improvements on MRM (IMRM)

As shown previously MRM is based on a power function relationship between DBR and KB. See Eq. (12). A sample plot is shown in Fig. 4 for the Bracknell site. The best-fit equation for the above data was found to be DBR = 0.1894. $(K_b)^{-1.1656}$ with the coefficient of determination equal to 0.95.

In an attempt to improve the estimation, the dataset was split into seven SF bands. These are detailed in Table 3. Since data in the range of $0 < SF < 0.4$ was closely populated, it was considered appropriate to classify the data into the above mentioned band rather than three consecutive bands, of 0.1 SF step.

It was found to be more productive to plot the above banded data in an ln(DBR) vs. NLKB, where NLKB $= -\ln(K_b)$ rather than DBR vs. K_b scatter plot as shown in Fig. 5. The relationship between NLKB and ln(DBR) is a 4th order polynomial in the form:

$$\ln(\text{DBR}) = a_0 + a_1(\text{NLKB}) + a_2(\text{NLKB})^2 + a_3(\text{NLKB})^3 + a_4(\text{NLKB})^4 \tag{19}$$

The above approach increases the R^2 of the regression. For Bracknell, the power function as defined in Eq. (12) yielded an $R^2 = 0.95$, while for all bands, lower SF

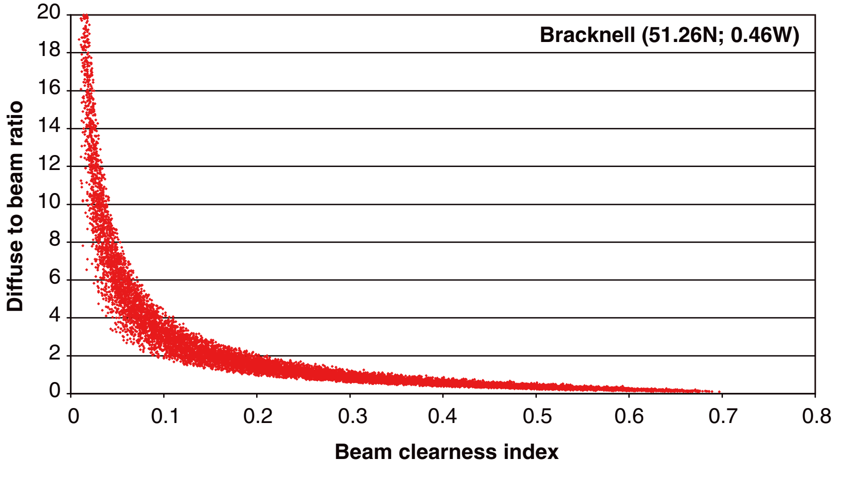


Fig. 4. Scatter plot of DBR vs. Kb for Bracknell (UK). Note that DBR and K_b are dimensionless.

Table 3
The seven SF bands used to split the datasets to improve the accuracy of the regressions

Band number	Sunshine fraction
1	$0 \le SF < 0.4$
2	$0.4 \le SF < 0.5$
3	$0.5 \le SF < 0.6$
4	$0.6 \le SF < 0.7$
5	$0.7 \le SF < 0.8$
6	$0.8 \le SF < 0.9$
7	$0.9 \le SF \le 1.0$

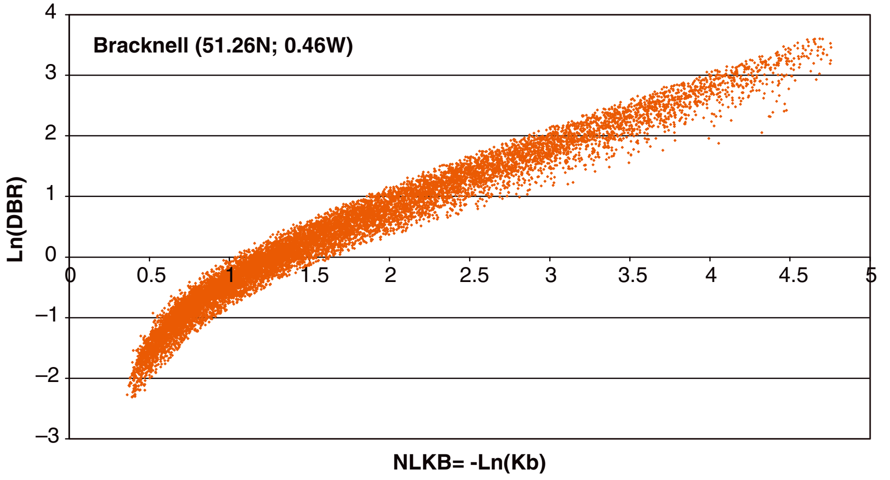


Fig. 5. Scatter plot of ln(DBR) vs. NLKB for Brackmell (UK). Note that ln(DBR) and NLKB are dimensionless.

limit = 0. 0 (exclusive) and upper SF limit = 1.0 (inclusive) the fourth order polynomial relationship between ln(DBR) and NLKB yielded an $R^2 = 0.976$. Thus, an improvement of 2.6% between the original and improved regression procedure was observed.

By plotting each data band's best fit line, it is noted that, where the bulk of the data is located, the best fit lines do not overlap each other and are in a particular order. Each specific data band has an $R^2 > 0.8$. It is, however, not possible to compare the bands based on the value of R^2 as the data population in each band is different.

The original MRM is modified in order to estimate the diffuse component of horizontal irradiance more accurately from the regressions between ln(DBR) and NLKB; thus improving the estimation accuracy of global horizontal irradiance.

The IMRM works in a similar manner as the original MRM by Muneer et al. [8–10]. For non-overcast skies, beam horizontal irradiance is a function of the extraterrestrial horizontal irradiance and the sunshine fraction and dampened by the transmittances due to Rayleigh, Mie mixed gases, ozone and water vapor transmittances. When processing the datasets, the improved MRM requires the same inputs as the original model, i.e. time stamp, DBT, WBT, P_{atm}, SF and the 14 coefficients for the transmittances equations. The improved model, however, requires 35 coefficients per site that are the result of the five coefficients of the 4th degree polynomial regression for each of the seven SF based data bands. Then by averaging each band regression coefficients for all 10 sites, a set of generalized regression coefficients we compiled, as shown in Table 4.

The regression lines in Fig. 6, representing each of the bands in the range $1.0 < -\ln(K_b) < 4$ were perfectly aligned, where the bulk of the data in each dataset lies. Due to scarcity of the data in the $-\ln(K_b) < 0.9$ and $-\ln(K_b) > 4.0$ bands, the correlation lines overlap and do not follow any particular order.

5.3. Validation of the IMRM

For validation purposes, the original MRM was used to estimate the global horizontal irradiance for 10 locations used. For each of 10 sites, the model was tuned by tuning the

Table 4
The 35 generalized coefficients needed for the IMRM based on Eq. (19) for the specified range of the data

SF Band	$-\ln(K_b)$,min	$-\ln(K_b)$,max	Regression Coefficients for Eq. (19)					
			a_0	a_1	a_2	a_3	a_4	R^2
1	0.000	5.000	−1.4786	0.3528	0.8191	−0.2744	0.0273	0.94
2	0.000	5.000	−2.6147	2.6782	−0.7084	0.1201	−0.0079	0.97
3	0.000	5.000	−2.8485	3.0903	−0.9961	0.2044	−0.0168	0.97
4	0.000	5.000	−2.8717	3.2533	−1.1234	0.2305	−0.0182	0.97
5	0.000	5.000	−2.8661	3.2627	−1.1995	0.2691	−0.0241	0.96
6	0.000	5.000	−2.6631	2.6301	−0.7453	0.1302	−0.0088	0.93
7	0.000	5.000	−2.53370	1.9348	−0.1349	−0.0751	0.0155	0.87

Note that the coefficient of determination for each of the 7 bands is above 0.85.

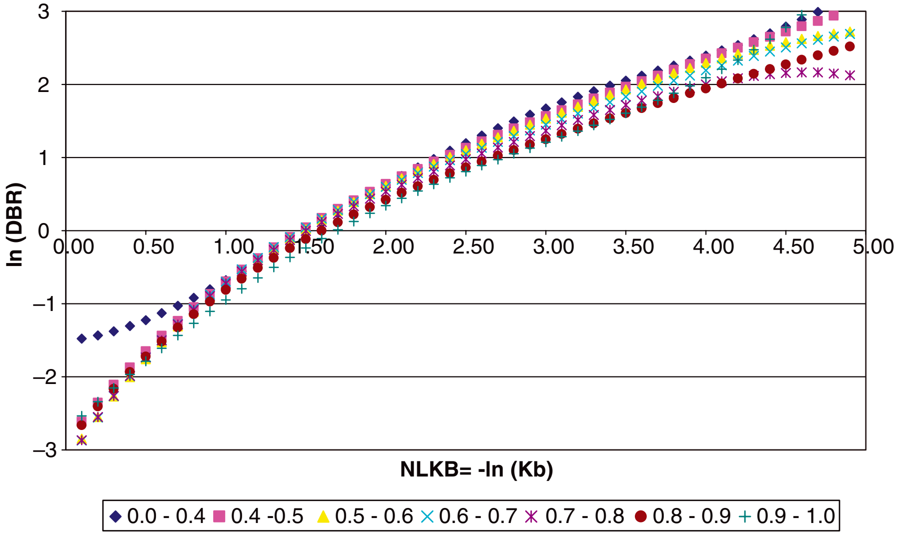


Fig. 6. Regression between ln(DBR) and NLKB for all seven SF bands. Note that ln(DBR) and NLKB are dimensionless.

14-transmittance coefficients, to suit the location and maximize the accuracy of the estimation. Since the original model has been validated, by Muneer et al. [7] using datasets from the UK and Japan, the same coefficients were used in the current validation for the UK and Japanese datasets. However, the validation of other seven sites two Spanish, four Indian and one Middle-eastern site, needed re-tuning for the 14-transmittance coefficients for a perfect regression relationship. Note that the MRM was only processed for data with solar altitude greater than 10°. Remark that for modeling purposes, the physical quality control test for solar altitude is more restrictive than the quality control procedure as presented in Section 3. A stricter test removes all suspected data that are located at sunrise and sunset.

The two models, original MRM and the proposed improved MRM were compared using statistical procedures/analysis. The coefficient of determination for the best fit line was calculated. Unfortunately R^2 is not a precise method of comparing the two models, as it is generally noted that the coefficient of determination increases with the amount of coefficients in the model. The original model relied on two coefficients in the regression between DBR and K_b, while the improved model relies on 35 coefficients in the regression between ln(DBR) and NLKB, thus the expected improvement in the coefficient of determination. To avoid misjudgment of the improved R^2 for new model two other statistical comparison methods: mean bias error (MBE) and root mean square error (RMSE) were used. The MBE provides an indication of the trend of the model, i.e. whether it has a tendency to under-estimate or over-estimate its modeled values.

Thus, for each location, two sets of three statistical parameters were employed, coefficient of determination of the best fit line between computed and measured global irradiance data, mean bias error and root mean square error for both the original MRM and the improved MRM. The 14 transmittances coefficients used for each site for the improved MRM were the same as for the original MRM. For the original MRM validation, the two regression coefficients, obtained in the correlation between DBR and K_b by the Muneer and Gul in their original validations [7–9], were used for all sites. For the improved MRM, the 35 generalized regression coefficients were used on all the sites.

5.4. Discussion of results

HSHDs and HDDs, comprising datasets from the 10 sites, were both processed by the original and the improved meteorological radiation models, by a case-by-case comparison, the improved model performed better with regards to reducing the amount of scatter and errors in the estimation process. This can be seen in Table 5 for HSHDs and in Table 6 for HDDs listing the statistical results for both the MRM and the IMRM.

The improvement can be visually observed in Figs. 7(a,b) for Bracknell HSHD and Figs. 8(a,b) for Mumbai HDD. Note that in Figs. 8(a,b) a considerable number of outliers are visible in the scatter plot. This is not an isolated case, as all HDD scatter plots for both the MRM and IMRM yielded considerable outliers. This is also confirmed by the results of the statistical analysis, in high MBE and RMSE values. This amount of scatter is caused by the nature of the HDDs. Single daily values of SF are assigned to hourly radiation values.

Table 5
Table shows validation results of HSHDs

Location	Model	R^2	MBE	RMSE
Bahrain	MRM	0.78	−47.1	170.5
	IMRM	0.75	12.3	151.9
Bracknell	MRM	0.81	−117.0	155.3
	IMRM	0.90	−19.4	68.3
Fukuoka	MRM	0.93	−67.4	96.9
	IMRM	0.92	30.5	76.2
Sapporo	MRM	0.71	−123.2	199.2
	IMRM	0.91	33.8	80.9

The MRM is the original model by Muneer et al. [7–9]. The IMRM is the proposed improved model.

Table 6
Table shows validation results of HDDs

Location	Model	R^2	MBE	RMSE
Chennai	MRM	0.41	134.9	254.5
	IMRM	0.58	−25.7	163.2
Gerona	MRM	0.37	−59.3	226.0
	IMRM	0.37	2.7	201.5
Madrid	MRM	0.63	−242.8	309.0
	IMRM	0.76	−121.9	171.1
Mumbai	MRM	0.26	−92.2	346.9
	IMRM	0.79	−15.1	114.8
New Delhi	MRM	0.78	61.1	119.3
	IMRM	0.91	74.8	94.0
Pune	MRM	0.68	−42.4	163.9
	IMRM	0.72	18.5	155.2

The MRM is the original model by Muneer et al. [7–9]. The IMRM is the proposed improved model.

If for a certain day, the SF value is given as 0.5; it does not necessarily indicate that every hour has an SF value of 0.5. In reality, a daily SF value of 0.5 might indicate that the morning hours could have been in clear-skies, while in the afternoon, sky was fully overcast. The outlier points, in the upper half of the scatter plots indicate that the daily SF values are low, sign of overcast-sky, while the high global horizontal irradiance values indicate clear-sky at the measured hours of the otherwise overcast-days and vice-versa for outliers in the lower half of the scatter plots.

Error histograms have also been plotted for each site from both model estimations, as can be seen in Fig. 9 examples for Bracknell and Madrid. It is another good indication of reduction in estimation error in the improved model as compared to the original model.

For further validation of the model, a residual examination was required, as discussed by Muneer [10]. This is done by a procedure that produces a graph of the residuals or errors d (the difference between the observed and calculated I_G) plotted against an independent variable n (the number of data points). In an ideal scenario, the plot of error has a horizontal shape. Further inspection of the errors plots suggests the model to be adequate enough, as shown in Fig. 10 plot of residuals for the Bracknell HSHD.

6. Cloud radiation model

6.1. Review of available cloud cover based radiation estimation models

Kasten and Czeplak [11] formulated equations Eqs. (20)–(22) for the estimation of solar irradiance based on cloud cover information. For their research, Kasten and Czeplak used 10 years (1964–1973) of continuous hourly data from Hamburg, Germany. They first proceeded by calculating the global horizontal irradiance under clear-sky conditions I_{GC}, then the global and diffuse horizontal irradiance. For the purpose of this analysis, the above mentioned model shall be referred to as M1a.

$$I_{GC} = 910 \sin \alpha - 30 \tag{20}$$

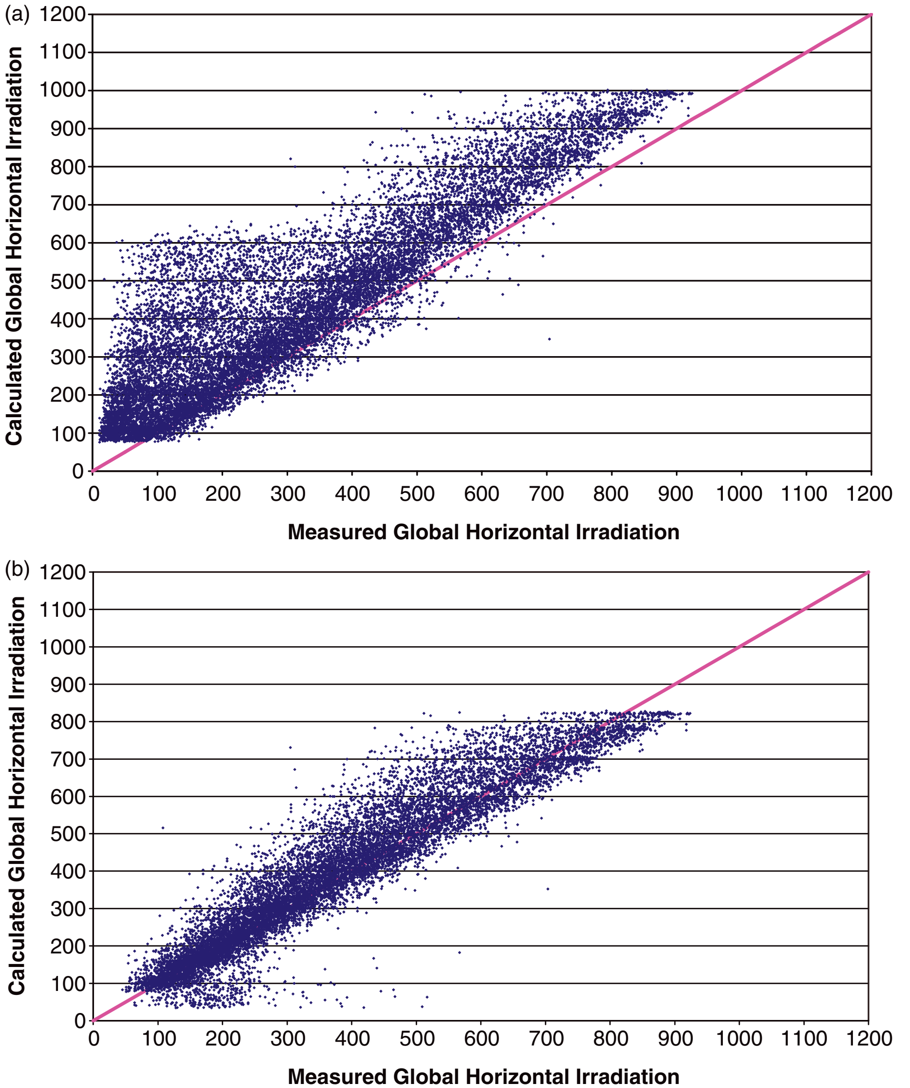


Fig. 7. Scatter plot of measured vs. calculated I_G for Bracknell using (a) the original MRM and (b) the proposed IMRM. Note that measured and calculated I_G values are in W/m^2.

$$I_G = I_{GC}\left(1 - 0.75(N/8)^{3.4}\right) \tag{21}$$

$$I_D = I_G\left(0.3 + 0.7(N/8)^2\right) \tag{22}$$

Gul et al. [12], and Muneer and Gul [13] have furthered the work of Kasten and Czeplak [11] to provide equations that can accommodate local coefficients for their datasets, as the original coefficients could not accurately estimate the irradiance in their analysis. The modified model shall be referred to as M1b in this analysis. The Kasten and Czeplak

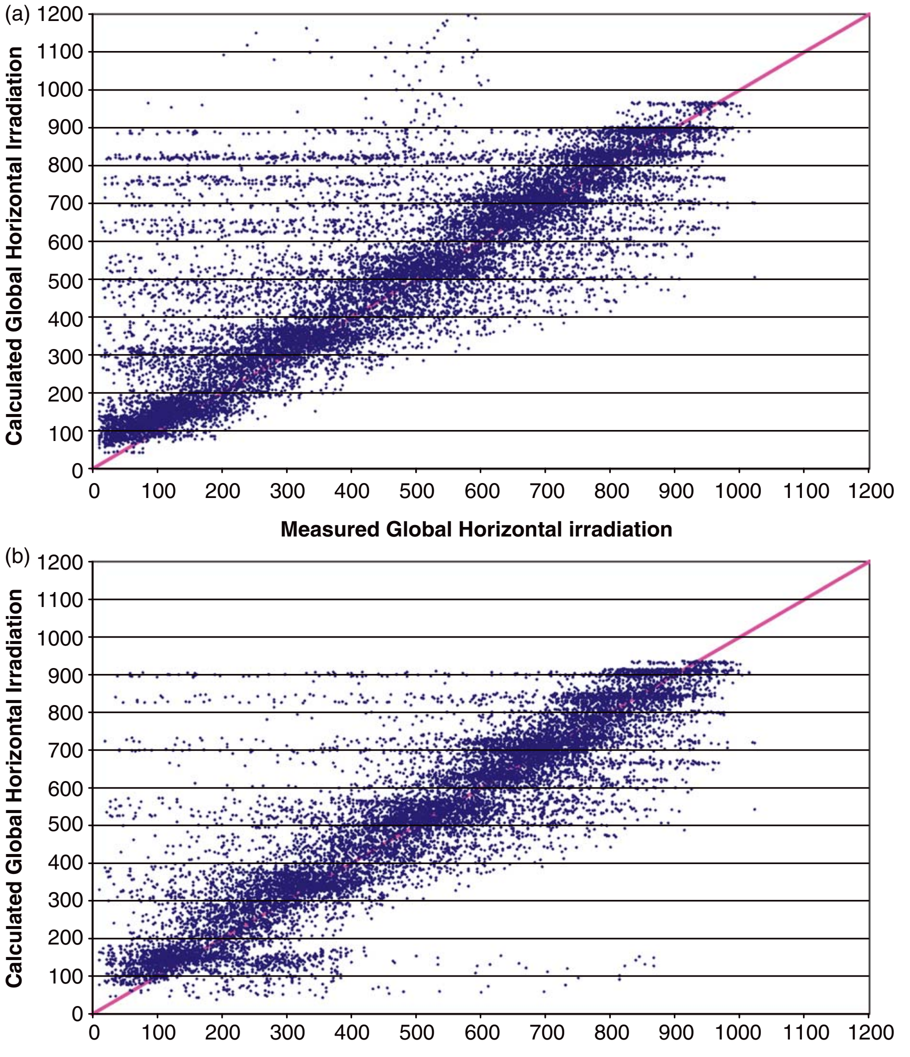


Fig. 8. Scatter plot of measured vs. calculated I_G for Mumbai using (a) the original MRM and (b) the proposed IMRM. Note that measured and calculated I_G values are in W/m^2. The outliers in the above figures can be attributed to the daily SF values assigned to hourly solar irradiance data.

equations Eqs. (20) and (21) have been modified while Eq. (22) remains the same in both models. Eqs. (20) and (21) become as follows:

$$I_{GC} = A_1 \sin\alpha - B_1 \tag{23}$$

$$I_G = I_{GC}\left(1 - C_1\left(N/8\right)^{D_1}\right) \tag{24}$$

Lam and Li [14] have explored the incorporation of multiple linear regressions between global irradiance and cloud cover involving solar altitude. Their model shall be referred to

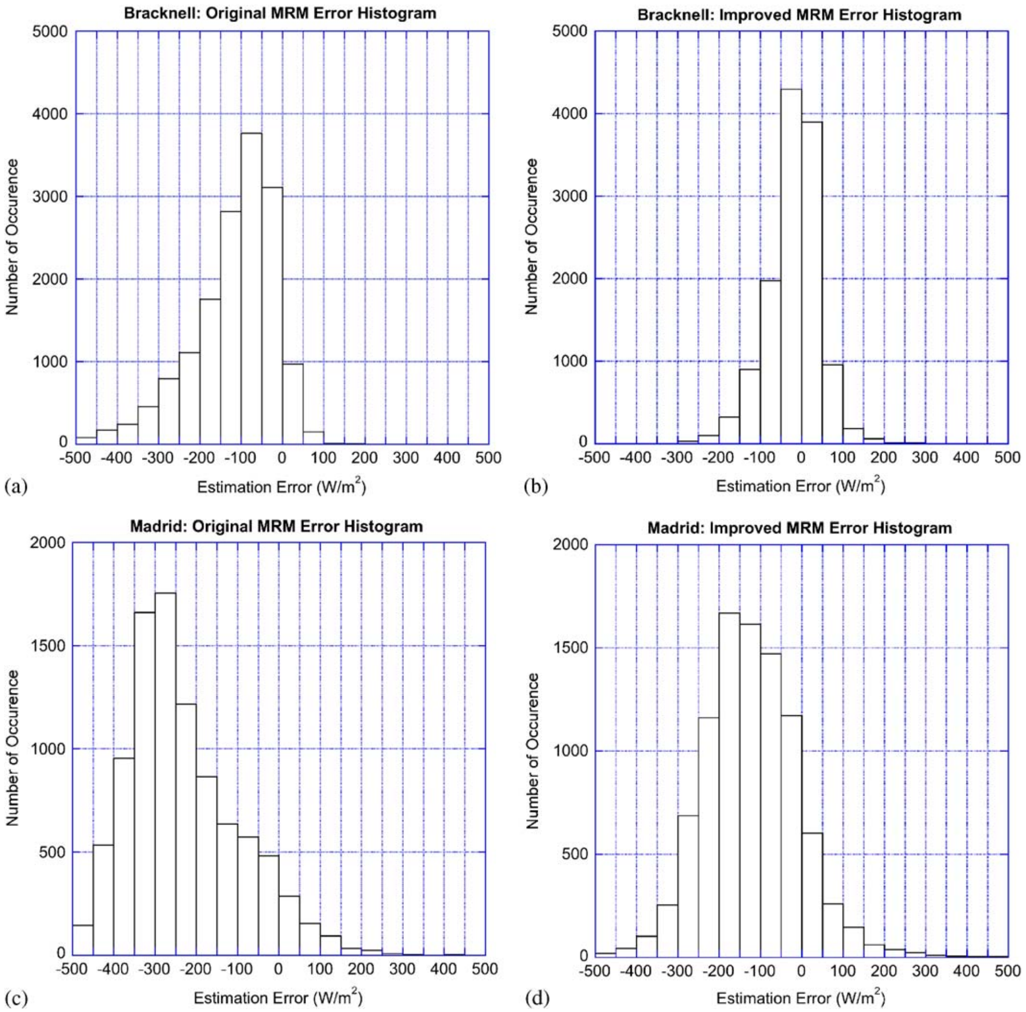


Fig. 9. The original and proposed improved MRM error histograms for the Bracknell and Madrid. Note that X axis represents radiation error in W/m^2 and Y axis is the number of occurrences.

as M2a in this analysis. The equations are given below,

$$I_G = 217 - 485(N/8) + 696 \sin \alpha \tag{25}$$

$$I_D = 30.5 - 62.9(N/8) + 294.7 \sin \alpha \tag{26}$$

The coefficients given in Eqs. (25) and (26) are in use for Hong Kong datasets, therefore, to accommodate for local coefficients for the datasets used in the analysis, the Lam and Li equations were modified as follows

$$I_G = A_2 - B_2(N/8) + C_2 \sin \alpha \tag{27}$$

$$I_D = D_2 -^{E_2} (N/8)^{+F_2 \sin \alpha} \tag{28}$$

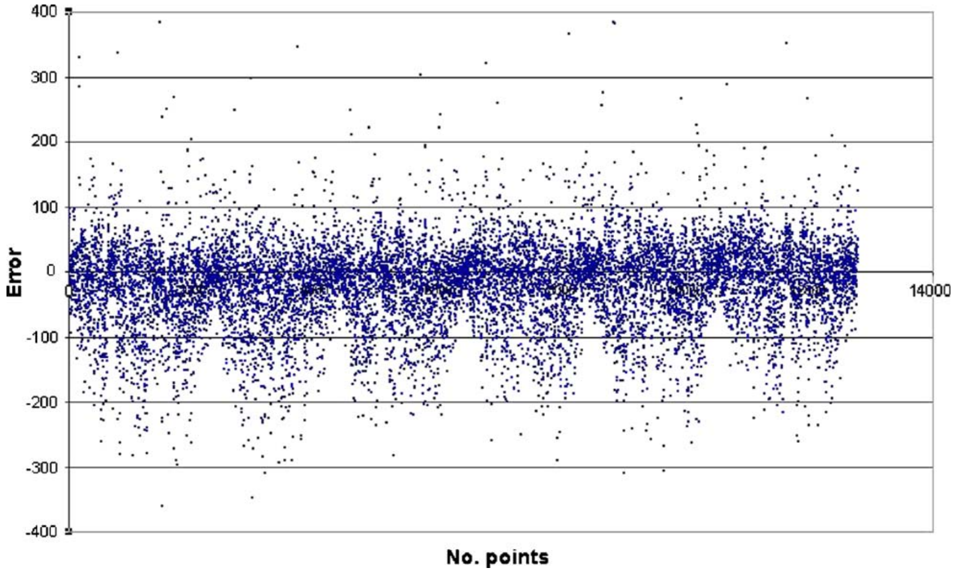


Fig. 10. Plot of residuals of the proposed improved MRM validation for the Bracknell HSHD site in the UK. Note that *Y* axis is the radiation error in W/m^2.

The modified Lam and Li equations shall be referred to as model M2b in the current analysis.

Note that in models M1a,b and M2a,b the beam horizontal irradiance is calculated by subtracting the diffuse component of the global horizontal irradiance, thus,

$$I_B = I_G - I_D \tag{29}$$

It is possible to identify weaknesses in the M1 models. Any errors in estimation in the first calculated component, I_{GC}, are transmitted to the second component calculated, I_G, and thereafter to I_D and I_B. This is due to the dependence of the component algorithms to each other. Therefore, if the user seeks I_B and/or I_D, then the errors found in the estimation are to be greater than for a user who seeks the estimation of I_G. Lam and Li [14] have identified this problem, and therefore, proposed two independent algorithms to estimate I_G and I_D. However, validation of models revealed flaws in M2a, and M2b. The Lam and Li models often result in negative estimation values. Moreover, the dependency problem applies in the estimation of I_B when using the Lam and Li derived models.

6.2. Analyzing cloud cover distribution

Cloud information is often used as an indicator of the sky conditions. Skies with no cloud are represented with a cloud cover index of 0 octas, and skies that are fully covered by cloud are therefore, represented by a cloud cover index of 8 octas. These represent the clear and overcast-skies, respectively. Sky conditions in between 1 and 7 octas are considered as mixed sky types

A good indication of the sky conditions can be obtained by analyzing cloud cover datasets. This is done by plotting a cumulative frequency of occurrence diagram. One such diagram is shown in Fig. 11 for Madrid.

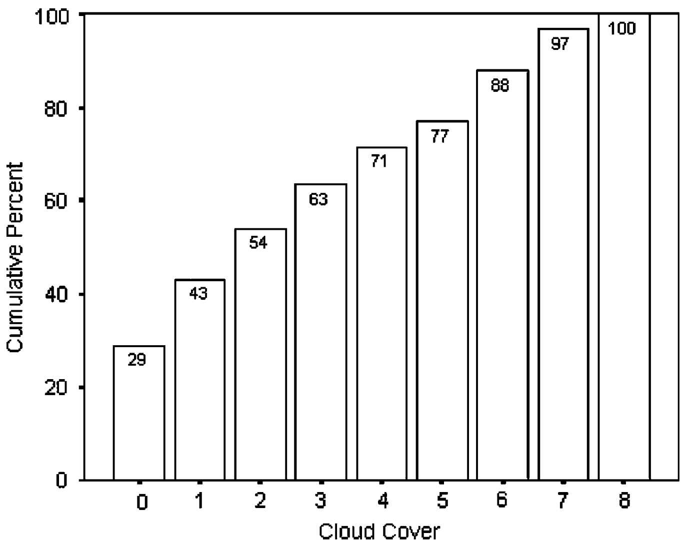


Fig. 11. Cumulative percentage frequency diagram of cloud cover for Madrid.

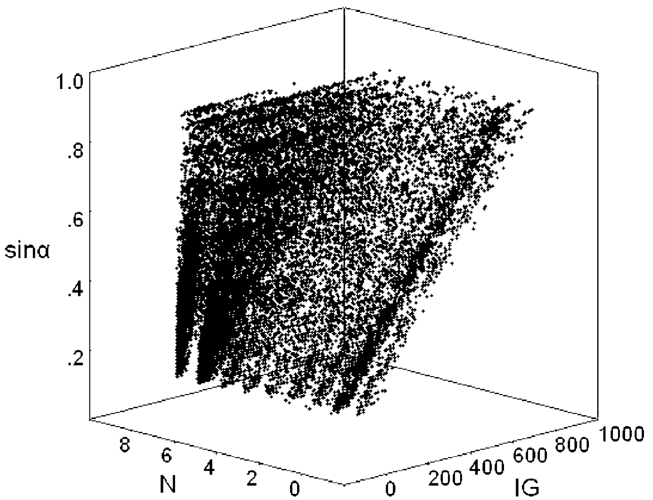


Fig. 12. Three-dimensional scatter plot of the sine of the solar altitude, cloud cover and I_G for Bracknell. Note that N in octas and I_G in W/m^2.

A relationship between the solar altitude, the cloud cover and the horizontal components of solar irradiance has often been used in the formulation of solar irradiance models that are based on cloud cover, as can be seen in models M1a, b and M2a, b, Eqs. (20)–(29). The relationship could be visually observed in Fig. 12, which presents a three dimensional scatter plot of cloud cover (N), sine of solar altitude and a horizontal solar irradiance component.

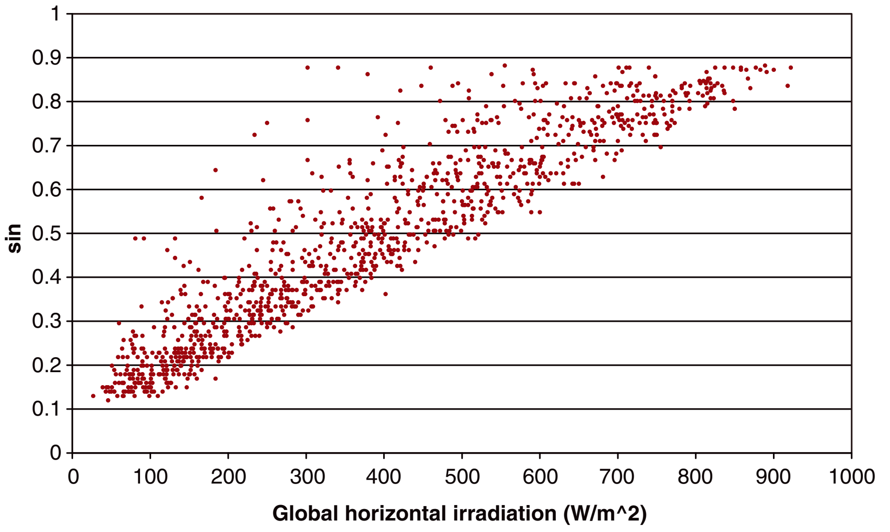


Fig. 13. Scatter plot of the sin α and I_G at cloud cover of 3 octas for Bracknell.

It was therefore, concluded that the solar irradiance components are dependent on the cloud cover and sinα,

$$I_i = f(\sin\alpha)f(N) \tag{30}$$

It was observed that there is a linear regression between the solar irradiance components and the sine of the solar altitude angle for each cloud cover band. This can be clearly observed from Fig. 12 and in Fig. 13. We thus conclude that Eq. (30) may be expanded thus,

$$f(\sin\alpha) = A\sin\alpha + B \tag{31}$$

Similarly the best regression representing the solar irradiance components and the cloud cover for each sine of solar altitude angle was found to be

$$f(N) = C + D\left(N/8\right)^E. \tag{32}$$

6.3. Improved multivariate models and hybridization

The proposed model, M3, was partly based on the Muneer and Gul [14] approach. The proposed formulations are:

$$I_G = (A_3\sin\alpha - B_3)\left(C_3 + D_3\left(N/8\right)^{E_3}\right) \tag{33}$$

$$I_B = (F_3\sin\alpha - G_3)\left(H_3 + I_3\left(N/8\right)^{J_3}\right) \tag{34}$$

$$I_D = I_G - I_B \tag{35}$$

As observed in Fig. 12, the cloud cover function is linear for cloud cover values under 4 octas for I_G. It was observed that there was no apparent linearity for I_B and I_D. Therefore, a hybrid bi-variate model, M4, has been developed. It is mathematically represented as, For CC$\leq$3,

$$I_G = (A_{4a} \sin \alpha - B_{4a})\left(C_{4a} + D_{4a}(N/8)\right) \tag{36}$$

and for CC$\geq$4,

$$I_G = (A_{4b} \sin \alpha - B_{4b})\left(C_{4b} + D_{4b}(N/8)^{E_{4b}}\right) \tag{37}$$

$$I_B = (F_4 \sin \alpha - G_4)\left(H_4 + I_4(N/8)^{J_4}\right) \tag{38}$$

Air mass was included in the regressions to calculate the horizontal solar irradiance, and thus creating a tri-variate model. This model will be referred to as M5, and is represented mathematically as,

$$I_G = (A_5 \sin \alpha - B_5)\left(C_5 + \left(D_5 + E_5 m + F_5 m^2\right)(N/8)^{G_5}\right) \tag{39}$$

$$I_B = (H_5 \sin \alpha - I_5)\left(J_5 + \left(K_5 + L_5 m + M_5 m^2\right)(N/8)^{O_5}\right) \tag{40}$$

Air mass (m) for each hour of the data was calculated using Kasten's formula [15], which provides an accuracy of 99.6% for zenith angles up to 89° [16].

$$m = \left[\sin \alpha + 0.50572(\alpha + 6.07995)^{1.6364}\right]^{-1} \tag{41}$$

Note that the only measured variable required for use with the above mentioned models is the cloud cover. The solar altitude angle and air mass are calculated independent variables.

6.4. Discussion of results

Due to the size of the databases involved, it was difficult in most cases to make a visual evaluation of all five models by examining the scatter plots of measured vs. calculated solar irradiance components. One may in a cursory manner observe that the M1b and M2b models perform better than M1a and M2a models. This indicates that using local coefficients significantly improves the performance of the models. It may also be noted that M1a and b perform better than M2a and b, respectively. However, it was difficult to compare the accuracy gains of M1a and b against those of M3, M4 and M5 from the scatter plots alone as this is only a qualitative method. Therefore, to quantitatively evaluate the performance of all models under discussion the following statistical indicators were employed. These indicators were the slope and coefficient of determination, R^2 of the best fit line between measured and calculated irradiance components. Furthermore, the mean bias error, MBE, and root mean square error, RMSE along with kurtosis and skewness of the error histograms related to the estimation of global, diffuse and beam irradiance were used.

Using the above mentioned statistical indicators an accuracy score, AS, was obtained for each model for I_G, I_B and I_D. Since the models, described above, estimate all three components of the horizontal solar irradiance, I_G, I_B and I_D then the maximum obtainable score for any model is 18, i.e. a total score of 6 per horizontal solar irradiance component.

Table 7
The ranges of the statistical indicators for the Aldergrove data evaluation

Statistical Indicator	I_G	I_D	I_B
Slope	0.71 to 0.84	0.50 to 0.83	0.55 to 0.76
R^2	0.70 to 0.81	0.61 to 0.74	0.55 to 0.73
MBE	−0.78 to 70.13	−7.00 to 33.56	−1.24 to 36.56
RMSE	84.12 to 130.57	52.69 to 72.93	70.32 to 104.55
Kurtosis	0.21 to 2.48	0.91 to 2.15	0.27 to 7.41
Skewness	−0.22 to 0.77	−0.28 to 0.63	0.12 to 1.92

However, obtaining the maximum score does not necessarily imply that the model is accurate and best performing; it only indicates that the model yields better results compared to the other models in the evaluation. It is, therefore, essential to take into consideration the statistical indicators. The range of the statistical indicators can be observed in Table 7 for Aldergrove.

For all the sites that were processed in this frame of work, it was found that the models that allow the use of locally tuned coefficients yielded better results than the models that use the original coefficients supplied by the authors. Therefore, for models M1 and M2, the 'b' variants performed better than the 'a' variants. This was also observed visually by inspecting both the scatter plots and the error histograms.

The statistical indicators and hence the AS indicate the superiority of the local coefficient variants of all models.

Models M2a and b, have a specific weaknesses; the algorithms used in the models cannot estimate the horizontal solar irradiance components of the available datasets to an acceptable degree of accuracy. It was noted, however, that the M2 model when fitted with local coefficients, M2b, does accurately estimate the global and diffuse horizontal irradiance for the clear-sky data. However, in all other cases the result was found to be very weak. In all cases, the M2 models yielded negative estimations, thus making the data obsolete since the irradiance cannot be negative. This can be observed in Fig. 14, scatter plot of measured vs. calculated I_G for Madrid using model M2b.

It was noted that model M1a performed exceptionally well in estimating I_G and I_D. This is due to the fact that Kasten and Czeplak had created the model using UK and German datasets. When comparing models M1a, and b, and models M2a, and b using the statistical indicators, models M2a, and b were found to be weaker their M1 counterparts.

Models M1a, M1b, M3, M4 and M5 all estimated I_G and I_D with good accuracy; however, M1a and M1b were weaker in the estimation of I_B. It was found that for models M3, M4 and M5, having separate equations to estimate I_D yielded little improvement on the procedures presented above.

Based on the AS results for the estimation of all three components of the horizontal solar irradiance, it was found that the proposed models performed better than the older models. This can be observed in Table 8, AS results for Bracknell.

Model M4 being a more complex version of model M3 does yield some improvements compared to M4, however, due to the increase in the number of coefficients used, the improvement is questionable. This is also the case for the tri-variate model M5, the addition of the air mass as a variable does not increase the accuracy of the estimation of the horizontal solar irradiance beyond model M4.

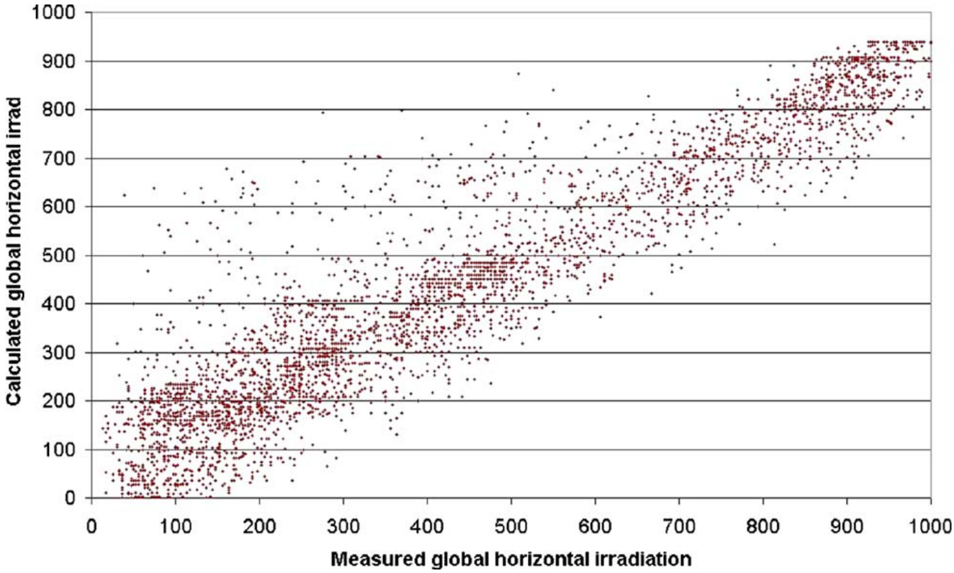


Fig. 14. Scatter plot of measured vs. calculated I_G for Madrid using model M2b.

Table 8
The accuracy score results of the CRM comparisons on Bracknell

No.	Model Name	Accuracy Score			
		IG (/6)	ID (/6)	IB (/6)	Total (/18)
M1a	Kasten-Czeplak	4.00	4.51	3.47	11.98
M1b	Muneer et al.	4.98	3.24	3.88	12.09
M2a	Lam-Li	2.65	2.40	2.77	7.83
M2b	Modified Lam-Li	4.06	3.70	3.45	11.21
M3	Proposed bi-variate	4.96	4.86	4.48	14.29
M4	Proposed hybrid bi-variate	4.89	4.94	4.46	14.30
M5	Proposed tri-variate	5.00	4.55	4.35	13.90

For all the datasets that were utilized in this paper, model M3 is the most accurate of the models compared. This is mainly due to the independence of the horizontal solar irradiance components. The proposed model M3 shares the same basic structure with M1a, b for the estimation of I_G, therefore, no particular improvement was observed, however, it was observed that M3 estimated I_B and hence I_D with greater accuracy than M1a, b or M2b. This can be observed visually in Fig. 15(a–i), scatter plots of measured vs. calculated I_G, I_D and I_B for models M1b, M2b and M3. Similarly, the improvement can also be observed in Fig. 16(a–i), I_G, I_D and I_B estimation error histograms for models M1b, M2b and M3. Both Figs. 15 and 16 are for the Madrid data. For the proposed model, M3, the coefficients used for each of the datasets that was used in this research are given in Table 9. It was observed that for Aldergrove, Bracknell, Gerona, Madrid and Mumbai, the coefficients were very similar and had very little variations. However, it was noted that

Chennai and Pune had different coefficients than the rest of the sites. This can also be observed in Fig. 17, distribution diagram of the coefficients used for model M3.

Barker [16] states that estimation of total cloud cover by real observations is subject to perspective errors and this causes inherent errors in the available datasets and is partly the reason for higher MBEs related to the estimation of I_G and I_D. Harrison and Coombes [17] noted that the weather observer generally overestimates clouds. In this regard Myers [18]

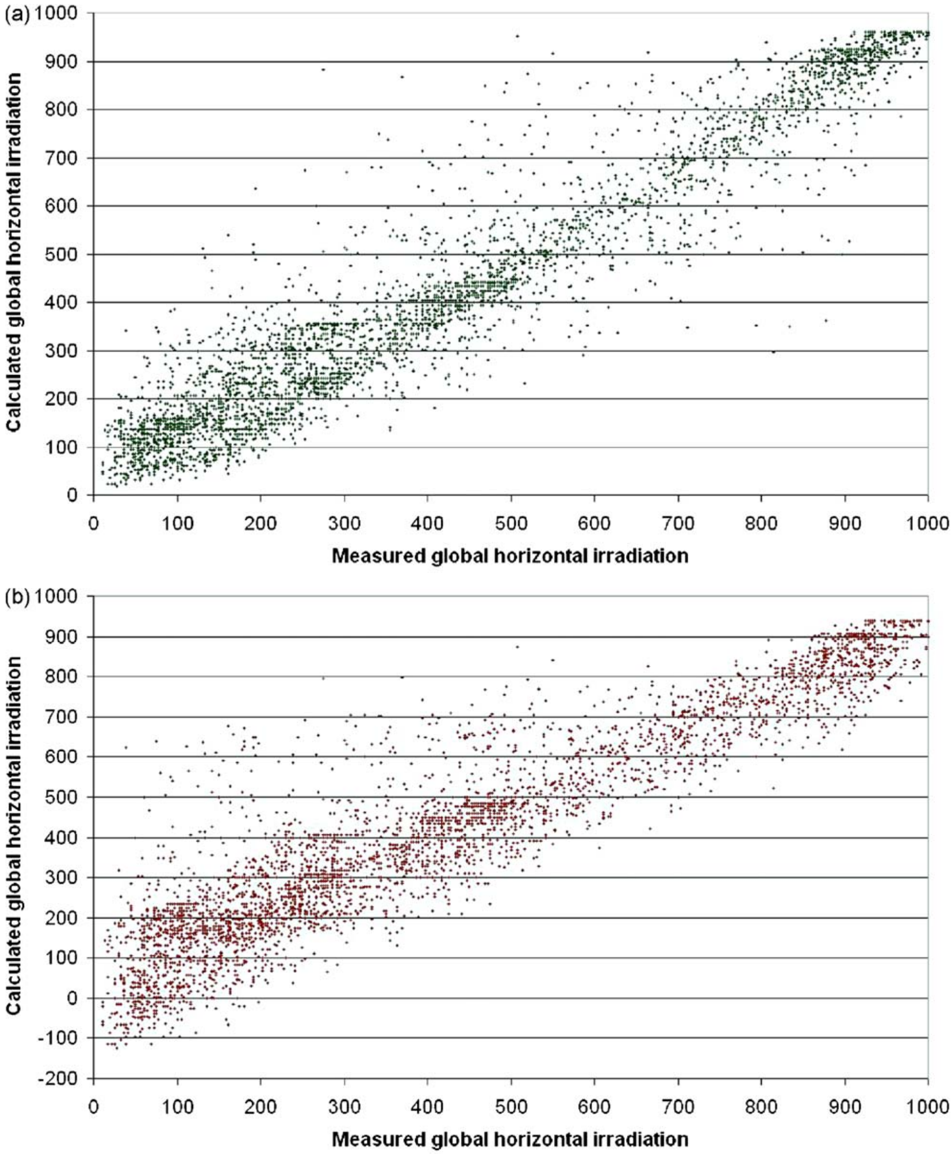


Fig. 15. Scatter plot of measured vs. calculated solar horizontal irradiance components, I_G, I_D and I_B for Madrid using model M1b, M2b and M3, respectively.

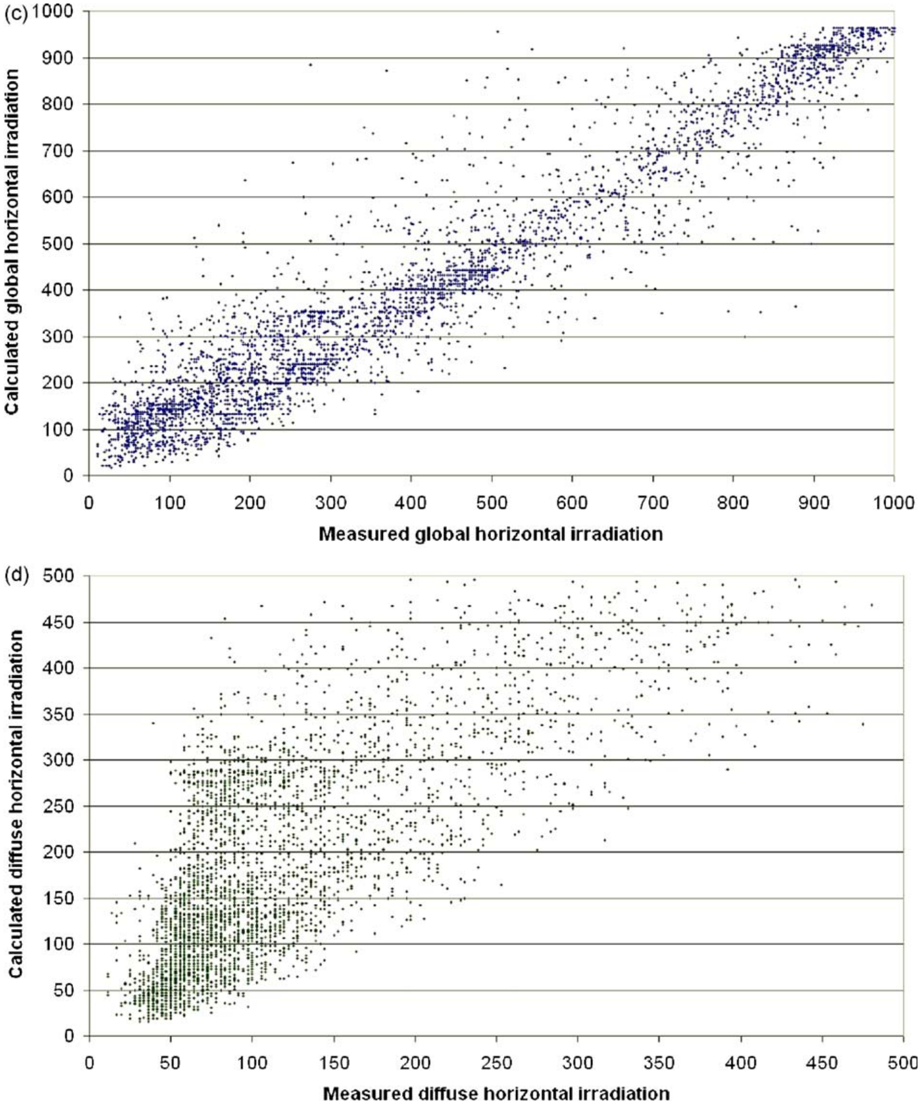


Fig. 15. (*Continued*)

remarks that the estimation models can only be proven as good as the data. In respect of his studies with uncertainties and error in solar radiation measurement instruments, Myers [18] states that absolute measurement uncertainties are of the order of 25–100 W/m^2 in hemispherical measured data. Brinsfield et al. [19] remark that an observer has a general tendency to underestimate the cloud cover under low overcast conditions and overestimate the cloud cover during high overcast conditions. Bennett [20] states that cloud cover explains less than 50% of insolation variance, while sunshine fraction per example explains between 70 and 85% of the insolation variance. The reason for the weakness of cloud cover

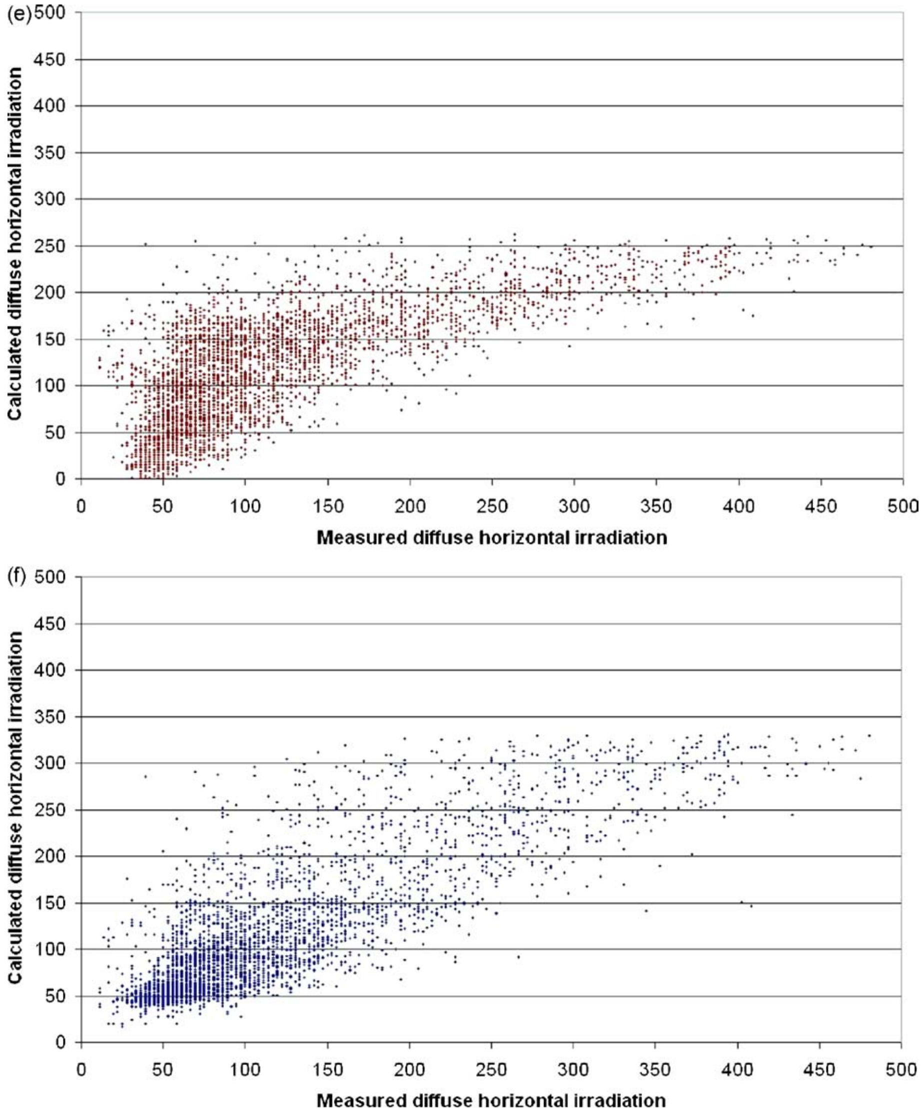
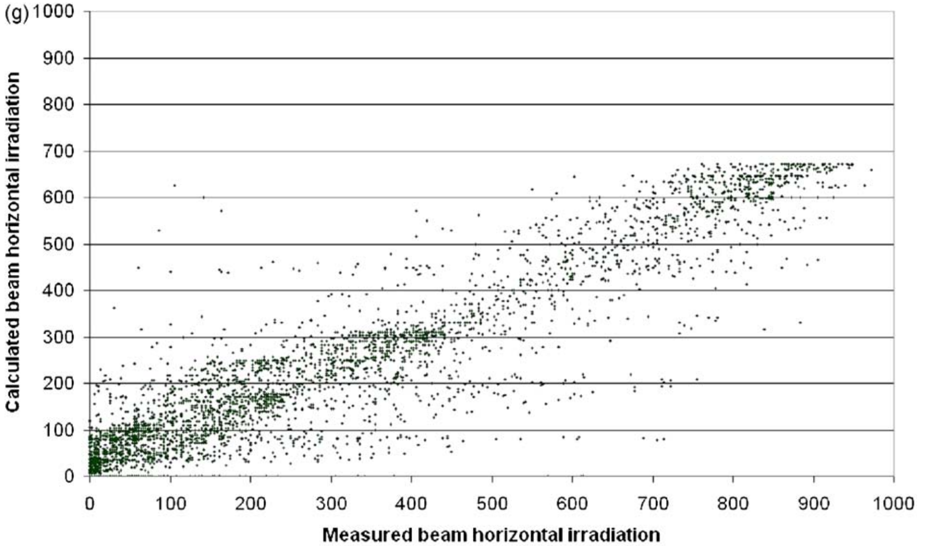


Fig. 15. (*Continued*)

is that it does not take into consideration the type and depth of the cloud. Cloud type varies immensely the amount of scattering and shading of the terrestrial solar radiation.

Coulson [21] classifies pyrheliometers and pyranometers in categories based on the uncertainties and errors of the measurements. A pyrheliometer is deemed 1st class, if the measurement errors compared to a reference pyrheliometer are in the range of $\pm 4\%$, while it is deemed 2nd class if the errors are in the range of $\pm 8\%$. Similarly for pyranometers, they are deemed 1st class, if the measurement errors compared to a reference pyranometer are in the range of $\pm 10\%$, 2nd class in the range of $\pm 25\%$, and 3rd class in the range of $\pm 32\%$.



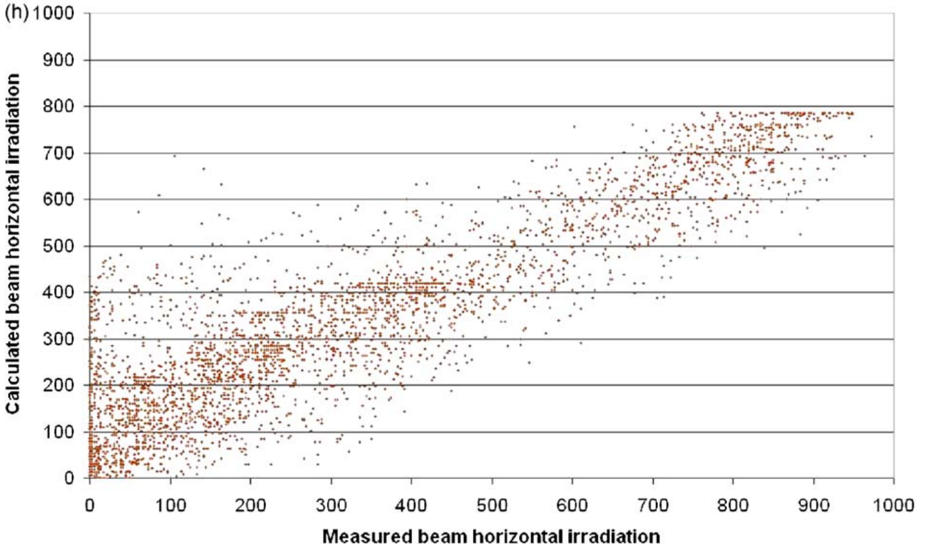


Fig. 15. (*Continued*)

With respect to this observation and referring to the work conducted by Muneer and Gul [13], the estimation of I_G and I_D datasets were divided in k_t bands to observe the patterns of estimation under overcast-, part overcast- and clear-skies. The data was divided in three k_t bands of $k_t \leq 0.2$, $0.2 < k_t < 0.8$ and $k_t \geq 0.8$, to respectively, represent the above sky conditions. MBE and RMSE were obtained to analyze the performance of the models. It was noted that all the models performed poorly under overcast-skies. This is due to the fact that k_t attains a value of 0.2 under overcast conditions, irrespective of the type of cloud. The models performed satisfactorily under the remainder of the sky conditions.

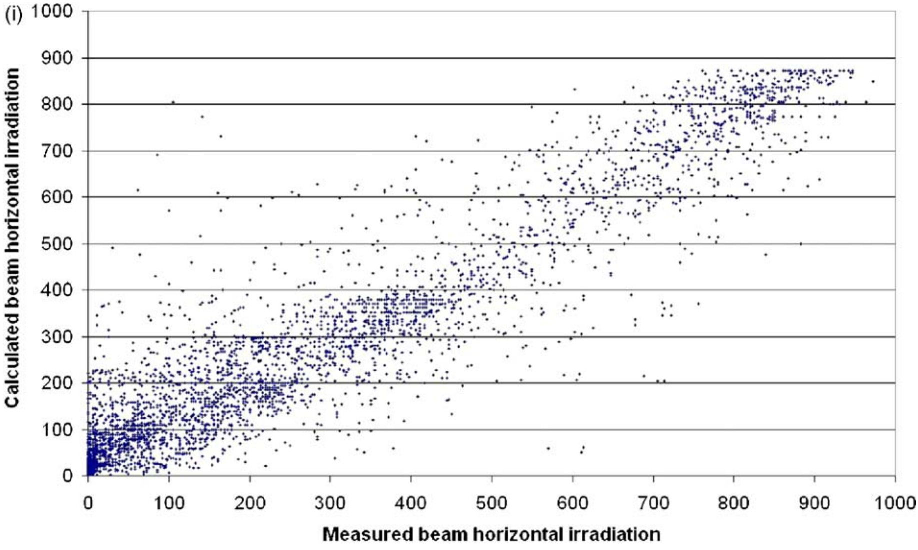


Fig. 15. (*Continued*)

7. Diffuse radiation model

7.1. Potential for improvement in the k–k_t regression model for the estimation of diffuse radiation

The very basic model, which shows a polynomial relationship between k and k_t, does not provide an accurate estimation of diffuse fraction of global radiation. There definitely remain other potential parameters that exert a significant influence on the diffuse fraction and consequently on diffuse radiation. A typical k–k_t plot shown in Fig. 18 displays a considerable scatter of data points for this chosen site (Bracknell, UK). Any suitable regression fit (linear, quadratic or cubic) does not and cannot account for the whole data. For example, a k_t value of 0.5 has observed values of $k = 0.35$ and 0.87, respectively, in clear and overcast regimes (marked by highlighted points in the figure). Whereas, cubic regression fit for the same data when k_t is 0.5, gives a value of $k = 0.61$, which overestimates the clear-sky observation by 74% and under estimates the corresponding overcast value by 30%. Likewise, quadratic regression fit yields a $k = 0.63$ for $k_t = 0.5$, yielding errors of a similar order of magnitude. Linear fit, as can be inspected visually, is not in the least better than the other two. For a typical data point, with $k_t = 0.5$, the diffuse horizontal radiation from quadratic regression model was calculated to be 141 W/m^2, whilst the actual observed value was as low as 79 W/m^2. This yields a large error of 78%, again emphasizing the fact, that diffuse radiation is very poorly predicted from k–k_t model. Furthermore, the value of $R^2 = 0.893$ obtained for the above quadratic k–k_t fit shows that 10.7% of the variation is unexplained (the definition of R^2 shall be explained later) and thus other factors have to be incorporated.

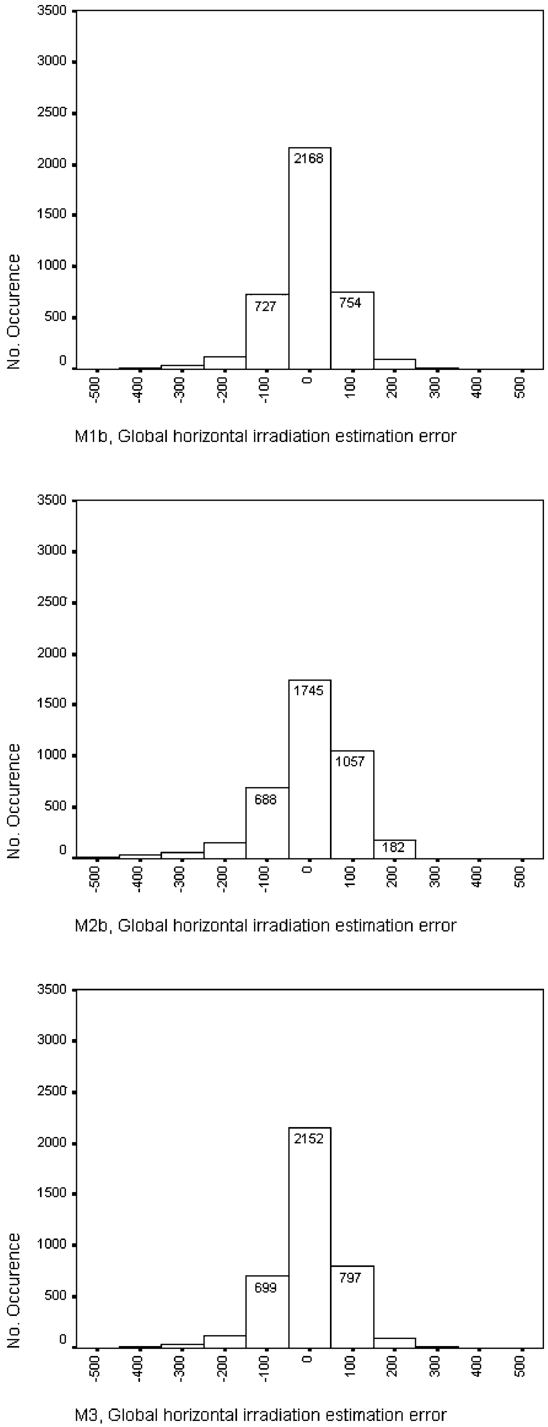
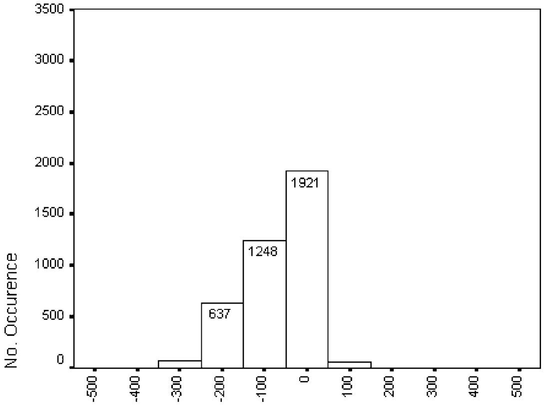
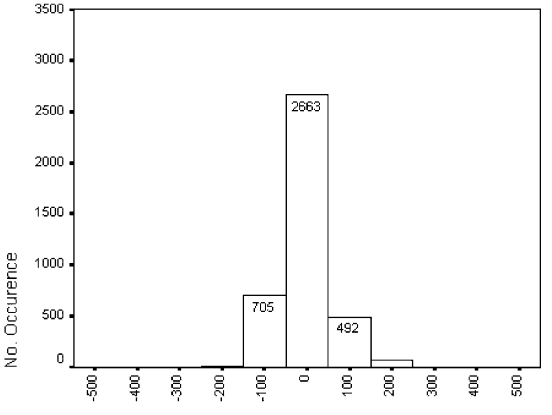


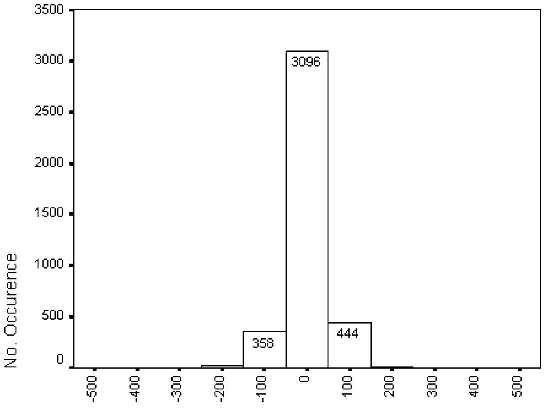
Fig. 16. Error histograms for solar horizontal irradiance components estimation, I_G, I_D and I_B for Madrid using model M1b, M2b and M3, respectively.



M1b, Diffuse horizontal irradiation estimation error

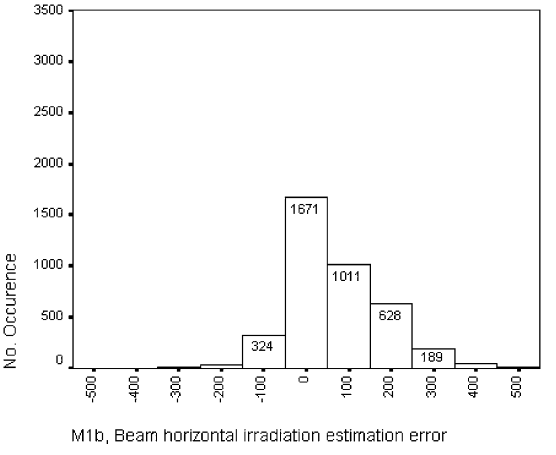


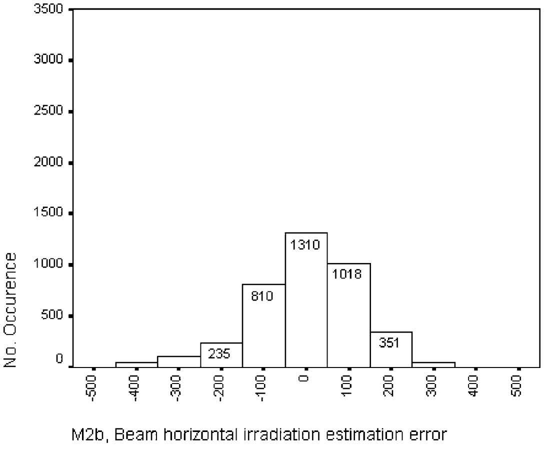
M2b, Diffuse horizontal irradiation estimation error



M3, Diffuse horizontal irradiation estimation error

Fig. 16. (*Continued*)





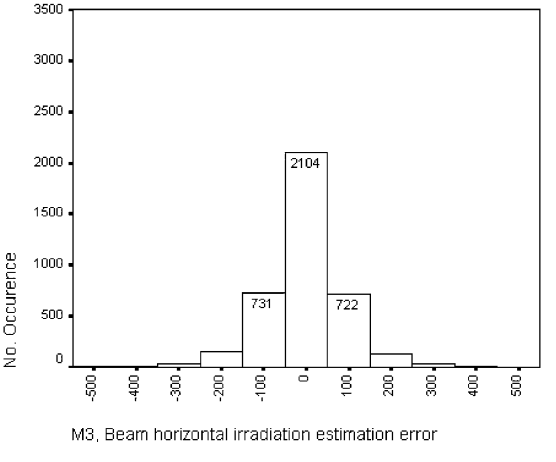


Fig. 16. (*Continued*)

Table 9
Values of coefficients used in model M3

Coefficient/ Location	Aldergrove	Bracknell	Chennai	Gerona	Madrid	Mumbai	Pune
A	1046	1046	1296	1046	1046	1046	1024
B	81	75	327	81	81	81	261
C	1.00	0.94	1.05	1.01	1.05	0.96	1.30
D	−0.73	10.68	10.38	−0.78	−0.65	−0.68	−0.74
E	3.17	4.02	2.09	2.37	3.06	4.55	5.25
F	998	995	1406	998	998	996	965
G	112	122	492	112	112	126	289
H	0.88	0.84	0.89	0.94	1.04	0.80	1.14
I	−0.90	−0.83	−0.81	−0.94	−0.96	−0.80	−1.09
J	1.47	1.73	1.52	1.53	1.81	2.45	2.70

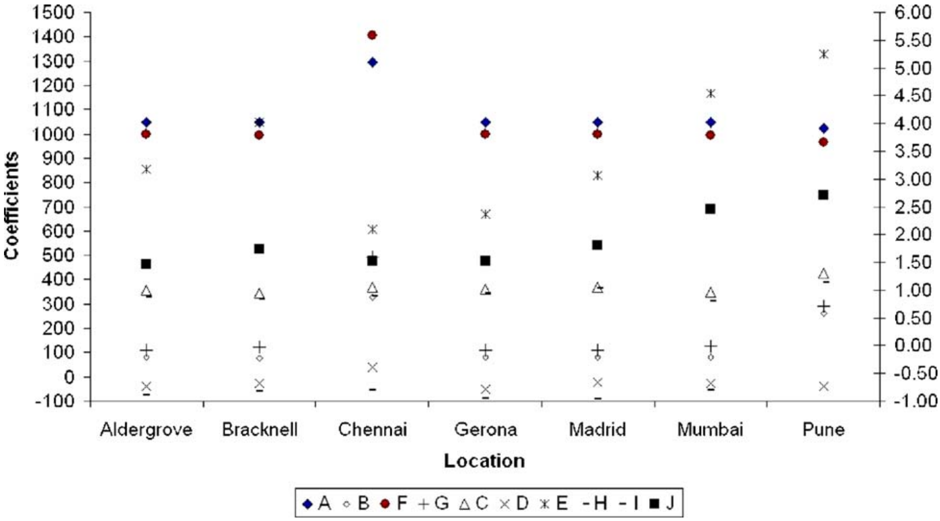


Fig. 17. Coefficients used for M3 per location. Coefficients A, B, F and G are plotted on the left *Y* axis; all other coefficients are plotted on the right *Y* axis.

7.2. Bearing of sunshine fraction, cloud cover and air mass on estimation of diffuse radiation

Various regressions have been developed in the past taking in account one or some of the following: atmospheric turbidity, atmospheric aerosols, relative humidity, ambient temperature, solar elevation, sunshine fraction (SF), air mass (m), cloud cover (CC) and albedo.

For the present work, daily SF (ratio of number of bright sunshine hours to day length) or hourly or sub-hourly sunshine fraction, as the case may be, CC and m were investigated as to explore their influence, if any, on the diffuse radiation estimation. While m can be calculated from the knowledge of solar altitude, measurement of SF and CC in contrast to the diffuse radiation is widely undertaken along with other meteorological data such as humidity and temperature. In UK itself, some 600 sites measure CC, out of which 230 sites

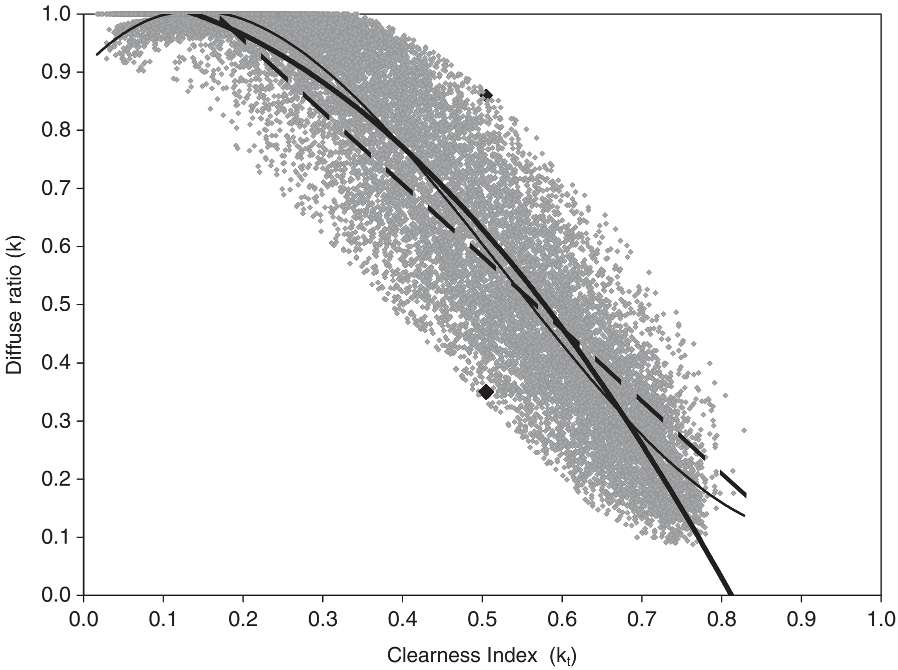


Fig. 18. k–k_t regression analysis for Bracknell quality controlled database. Dashed curve represents linear fit ($R^2 = 0.855$), thick solid line is for quadratic polynomial fit ($R^2 = 0.893$) and thin solid line for cubic polynomial fit ($R^2 = 0.903$).

measure sunshine duration while diffuse radiation is measured at relatively fewer sites. When the sky is completely covered by clouds (8 octas of cloud cover), the sunshine fraction is zero, and the radiation received on the earth's surface is completely diffuse in nature. Likewise, for completely clear-skies, i.e. with nil cloud cover and sunshine fraction equal to unity, the global radiation chiefly comprises of beam component. Hence, it is logical to say that both SF and CC can potentially play a significant role in the determination of the diffuse fraction of global radiation. Regarding, the third parameter presently under investigation, i.e. m it is known that the scattering effects of greater air mass render considerable portion of global radiation as diffuse component at lower solar altitudes [22]. This implies that for a given value of k_t an increase in m can lead to a corresponding increase in k. We preferred air mass to solar altitude, because the former is a more appropriate parameter for the characterization of atmosphere, e.g. turbidity and aerosol loading.

The following section investigates the qualitative effect of SF, CC and m on k–k_t relationship using Bracknell as a sample site. In the present analysis, k is taken as a quadratic function of k_t. This polynomial relationship was chosen to provide a simple yet a widely applicable platform in order to explore the other factors which might lead to more accurate estimates of diffuse radiation as compared to using k–k_t models alone.

7.2.1. Regressions under sunshine fraction (SF) bands

The available data for each site is divided into appropriate bands of SF, which ranging from 0 to 1. For each band of SF, a quadratic regression curve is fitted for the data in that

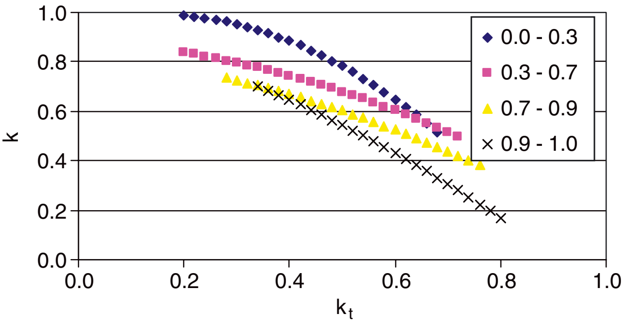


Fig. 19. k vs. k_t regression curves for different SF bands.

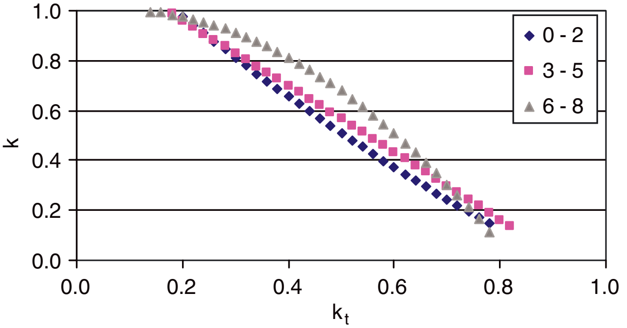


Fig. 20. k vs. k_t regression curves for different CC bands.

band. Fig. 19 shows one such plot. It is quite evident that, even though the bands are very close, their respective regression lines are distinct and separate. Logically, the higher the SF value, the lower would be the value of k for same k_t. In other words, higher SF bands are expected to fall in the lower region of k–k_t envelope and lower bands in the upper region. The fact that the overall generalized trend is in accordance with the trend that is theoretically predicted for all sites irrespective of their geographical/topographical variations, suggests that SF has a strong bearing on k–k_t relationships and can be used as an effective tool for diffuse radiation estimation. It was also reported by Gopinathan and Soler [23] in their study based on monthly mean daily values that, when both clearness index and relative sunshine duration are used together in multiple regressions the estimated values of diffuse radiation is better than when they are used separately.

7.2.2. Regressions under cloud cover (CC) bands

A similar analysis, as that for SF, was carried out for those six databases for which CC was available. For this part of the work, the data was split in three CC bands for all sites. Again, it can be interpreted that the regressions for these categorized data are distinct and follow a unique order for all the sites (Fig. 20). Higher CC band defines the lowermost curve, while lower values represent the uppermost curve and in the middle lays the intermediate CC band. Like SF, CC banded regression curves also deviate from the general trend towards the two extremes of overcast- and clear-sky conditions. Since the CC bands

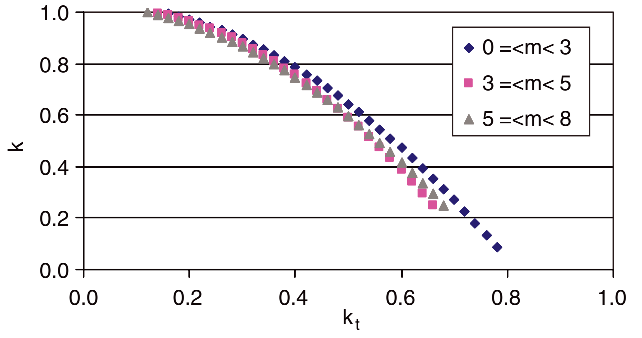


Fig. 21. k vs. k_t regression curves for different m bands.

are wider than the SF, they are expected to be more discrete. However, if the CC bands were only a unit (1 octa) wide the regression lines would not have been as distinct as that of SF. Conclusively, CC may not have as strong a potential as SF, but it does hold significance as one of the influential parameters for diffuse radiation modeling.

7.2.3. Regressions under air mass (m) bands

In this part of the investigation, k–k_t regressions were developed for different bands of air mass, varying from a minimum of 1 to a maximum of 8. The data sets were divided in three air mass bands. No specific trend for the k–k_t regressions under banded m data was observed (Fig. 21). The regression curves were completely random and overlapped each other. This can well be explained from the fact that same air mass values can lead to disparate sky conditions. Also, the effect of air mass is more apparent for clear than for overcast-skies. Thus, it can be concluded that air mass is a comparatively weak parameter for diffuse radiation estimation if used alone alongside k_t.

7.3. Modeling of diffuse radiation: the present approach

7.3.1. Effect of clearness index on diffuse ratio

It is a well-known fact that clearness index (k_t) strongly influences the diffuse ratio (k), and hence it is an integral parameter for the estimation of diffuse radiation. For that reason, we start from the basic k–k_t model and then retain k_t as the main regressor as we go along developing models with different combination sets of all the four parameters, namely, k_t, SF, CC and m. In the present analysis, we explored k–k_t regression equations with k as a polynomial function of k_t up to 3rd order. The quadratic fit was found to be the most adequate for all the nine sites. A plot of the data of Bracknell with three regression fits, as an example, is shown in Fig. 18. Although, as expected, the coefficient of determination increases with an increase in complexity of equation, the increase from quadratic to cubic regression model is not as significant as the improvement from a linear to quadratic fit. Moreover, the R^2 does not always measure the appropriateness of the model, since it can be artificially inflated by adding higher order polynomial terms in the regression Eq. (25). Also, note that if a cubic or quartic relationship was chosen it would have led to an unreasonable number of coefficients due to the interactions between the four parameters of k_t, SF, CC and m. Therefore, we select the second-degree polynomial k–k_t

relationship as the optimum and use it throughout our analysis for the sake of simplicity as well as uniformity.

$$k = a_{10} + a_{11}k_t + a_{12}k_t^2 \tag{42}$$

7.3.2. Effect of sunshine fraction, cloud cover and air mass on diffuse ratio: individual basis with clearness index as the common parameter

The following regression equation is used to investigate the effect of hourly clearness index and sunshine fraction or cloud cover or air mass on the hourly diffuse ratio.

$$k = \left(a_{i0} + a_{i1}X + a_{i2}X^2\right) + \left(b_{i0} + b_{i1}X + b_{i2}X^2\right)k_t + \left(c_{i0} + c_{i1}X + c_{i2}X^2\right)k_t^2 \tag{43}$$

where, $i = 2-4$ and X is SF, CC or m, respectively.

It is worthwhile to note here, that although the expression given above is the quadratic form of regression, two models based on both quadratic as well as linear function, were developed for each of the parameters.

7.3.3. Effect of clearness index, sunshine fraction, cloud cover and air mass on diffuse ratio: combined basis

Under this section, regression equations involving a combination of two variables and later on a tripartite variable combination, besides k_t, were explored. All the possible combinations were applied, not neglecting of course, the linear and quadratic models for each parameter.

$$k = \left(a_{50} + a_{51}X + a_{52}X^2 + a_{53}Y + a_{54}Y^2\right) + \left(b_{50} + b_{51}X + b_{52}X^2 + b_{53}Y + b_{54}Y^2\right)k_t + \left(c_{50} + c_{51}X + c_{52}X^2 + c_{53}Y + c_{54}Y^2\right)k_t^2 \tag{44}$$

where, $i = 5-7$, X and Y are SF and CC, or SF and m, or CC and m, respectively.

The regression equation proposed to evaluate the effect of the independent variables k_t, SF, CC and m taken all together is represented in the form given by Eq. (45),

$$k = \left(a_{80} + a_{81}\mathrm{SF} + a_{82}\mathrm{SF}^2 + a_{83}\mathrm{CC} + a_{84}\mathrm{CC}^2 + a_{85}m + a_{86}m^2\right) + \left(b_{80} + b_{81}\mathrm{SF} + b_{82}\mathrm{SF}^2 + b_{83}\mathrm{CC} + b_{84}\mathrm{CC}^2 + b_{85}m + b_{86}m^2\right)k_t + \left(c_{80} + c_{81}\mathrm{SF} + c_{82}\mathrm{SF}^2 + c_{83}\mathrm{CC} + c_{84}\mathrm{CC}^2 + c_{85}m + c_{86}m^2\right)k_t^2 \tag{45}$$

Regression analysis for the models given by Eqs. (42)–(45) was carried out for each location. Notice that although regression models presented here are in quadratic form, a linear fit was also tried and the better of the two was selected for each parameter of each site. That is to say, if a linear model in SF was found to deliver comparable or better accuracy than a corresponding quadratic one, the former model was chosen. To avoid repetition the equations with same parameters, but linear fit are not presented here. More so, quadratic form gives the reader an explicit idea of the scope of a given expression.

7.4. Evaluation of developed models

7.4.1. D_m vs. D_c plots

As a next step, the calculated diffuse radiation is plotted against the measured value for all the nine sites. Due to brevity of space, only four such plots are presented here one for each region Fig. 22.

The improvement from the conventional k–k_t model is evident with the addition of a single independent parameter SF. However, the model with all the parameters taken together shows a further improvement as can be seen from the significantly less scatter and their proximity to the one-to-one line in D_c vs. D_m plots. This trend was observed indisputably throughout all the sites. Hence, the model constituted by the pre-investigated optimum variable-relationships gives the most accurate estimation of diffuse radiation in the present context of the sites under study.

7.4.2. Error histograms

For each data point under investigation the calculated value of diffuse radiation was obtained form the above 'k' models. This calculated value was then subtracted from observed diffuse radiation. The errors thus generated, having the same units as diffuse radiation, i.e. W/m^2, were plotted against their frequency of occurrence for each model. In view of economy of length of publication, such error histograms for eight short listed

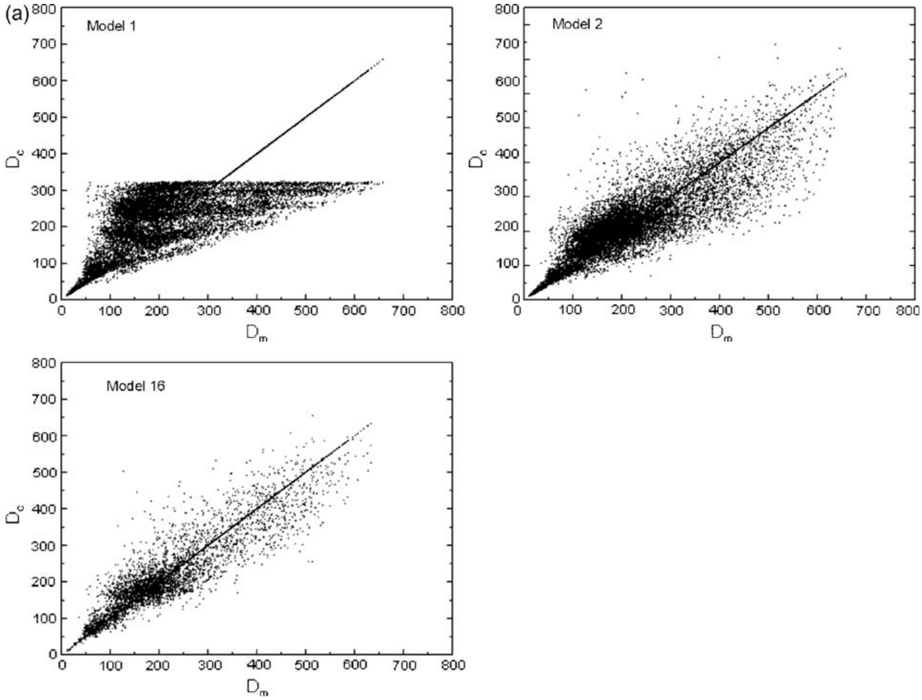


Fig. 22. The above plots show a significant improvement in the estimation of diffuse radiation from the basic k–k_t model through an intermediate (k–k_t, SF) to the eventual models selected as optimum. (a) Pune, (b) Fukuoka, (c) Gerona, (d) Bracknell.

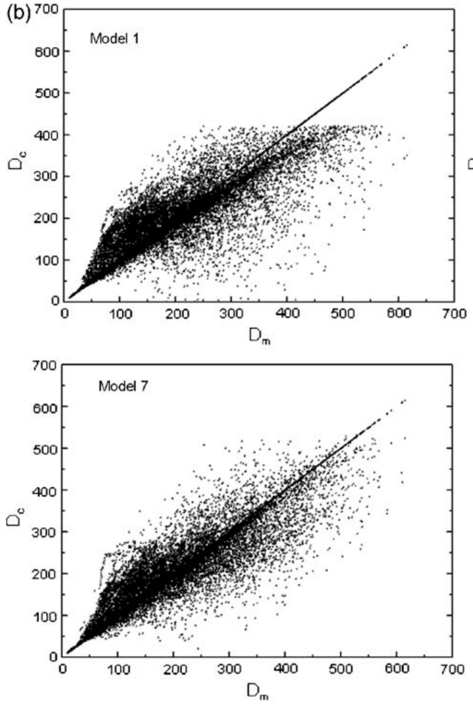


Fig. 22. (*Continued*)

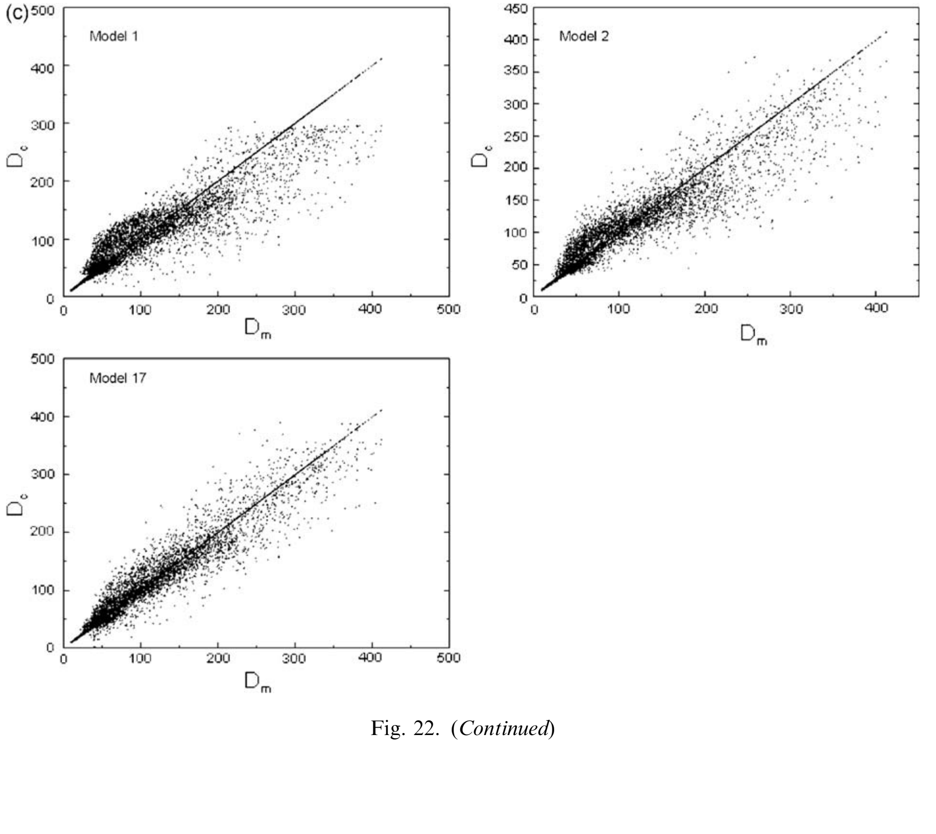


Fig. 22. (*Continued*)

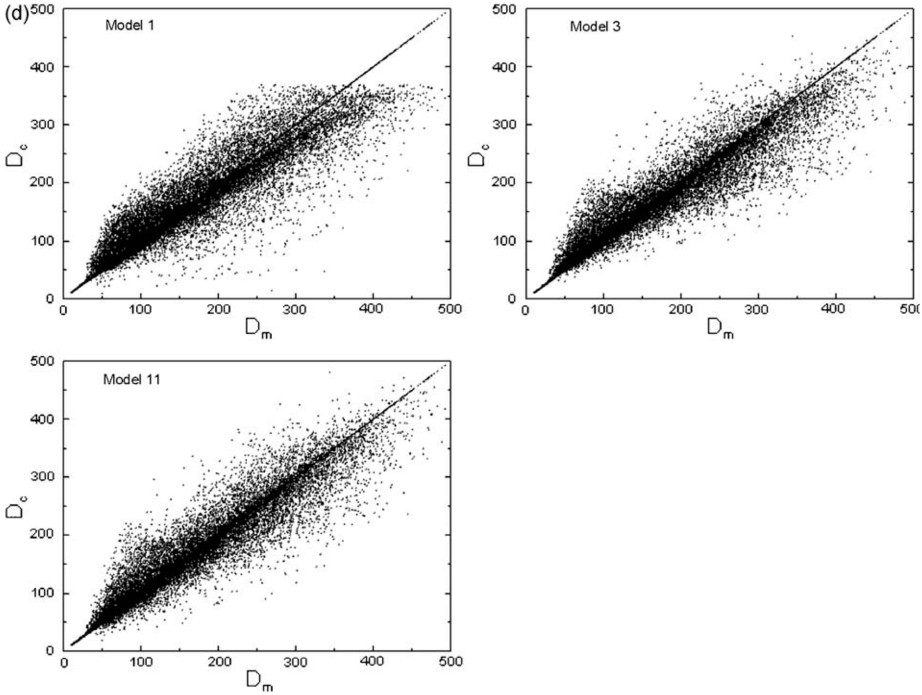


Fig. 22. (*Continued*)

models of only two sites, Mumbai and Gerona, are presented here Figs. 23 and 24. It is evident from the latter figures that with an increase in number of parameters error histograms improve, i.e. the peakedness of the error distribution increase around the zero error bands. This was a unanimous trend for all the sites. The simplistic k–k_t model or 'Model 1' has the widest error distribution, thus reflecting poor model performance. There is a significant increase in performance (more than 40%) from Models 1–8, in context of error frequency in the range of -10 to $+\mathrm{C}10\,\mathrm{W/m^2}$ for Mumbai and -5 to $+5\,\mathrm{W/m^2}$ for Gerona as shown in Figs. 22 and 23, respectively. Although, the error histograms for other sites are not presented here but other indicators, discussed in the following section, convey the error statistics of their models quite adequately.

7.5. Region-wise recommendation of optimum models

It was found that among all the models, the composite model involving all parameters provides the most accurate estimation of diffuse radiation. Analyzing the models on a comparative scale, it was also found that a single model could more than adequately estimate the diffuse radiation for the locations within a given region, which do not differ widely in their local climate. However, it is important to note here that although air mass is a calculated parameter and does not require actual measurements; both SF and CC are not always recorded by all meteorological stations. For instance, under the present context, there was no information available on CC for New Delhi and the two Japanese sites. Therefore, this emphasizes the need to recommend the optimum models also keeping in

view the availability of parameters. Three types of models were recommended for each region (highlighted model-types in Tables 10–13) based on the overall assessment of Accuracy Scores for each site within that region.

1. $k = f(k_t, \mathrm{SF}, m)$—for locations which provide information only on global radiation and sun shine duration.

For India:

$$k = \left(0.951 + 0.117\mathrm{SF} - 0.061m + 0.009m^2\right) + \left(2.268 - 2.277\mathrm{SF} - 1.335m + 0.207m^2\right)k_t + \left(-4.014 + 1.835\mathrm{SF} + 2.542m - 0.367m^2\right)k_t^2$$

For Japan:

$$k = \left(0.931 - 0.397\mathrm{SF} + 0.758\mathrm{SF}^2 + 0.009m\right) + \left(0.721 - 2.252\mathrm{SF} - 0.354\mathrm{SF}^2 - 0.126m\right)k_t + \left(-1.834 + 5.182\mathrm{SF} - 2.857\mathrm{SF}^2 + 0.271m\right)k_t^2$$

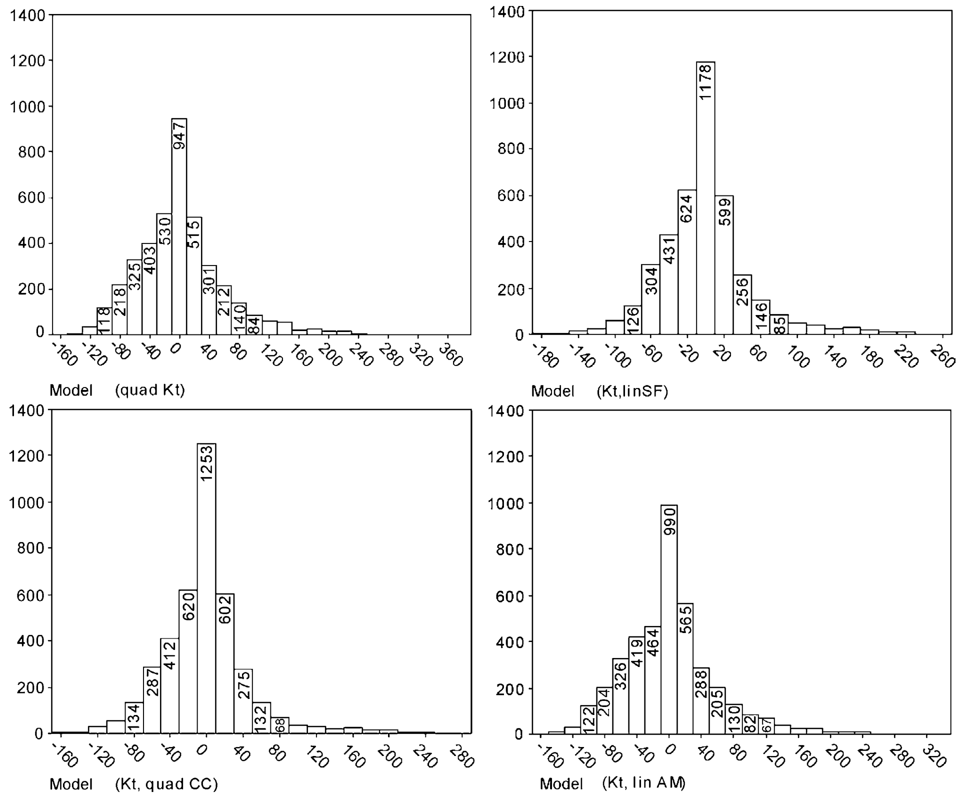


Fig. 23. Error histograms of calculated diffuse radiation of selectively chosen eight models for Mumbai.

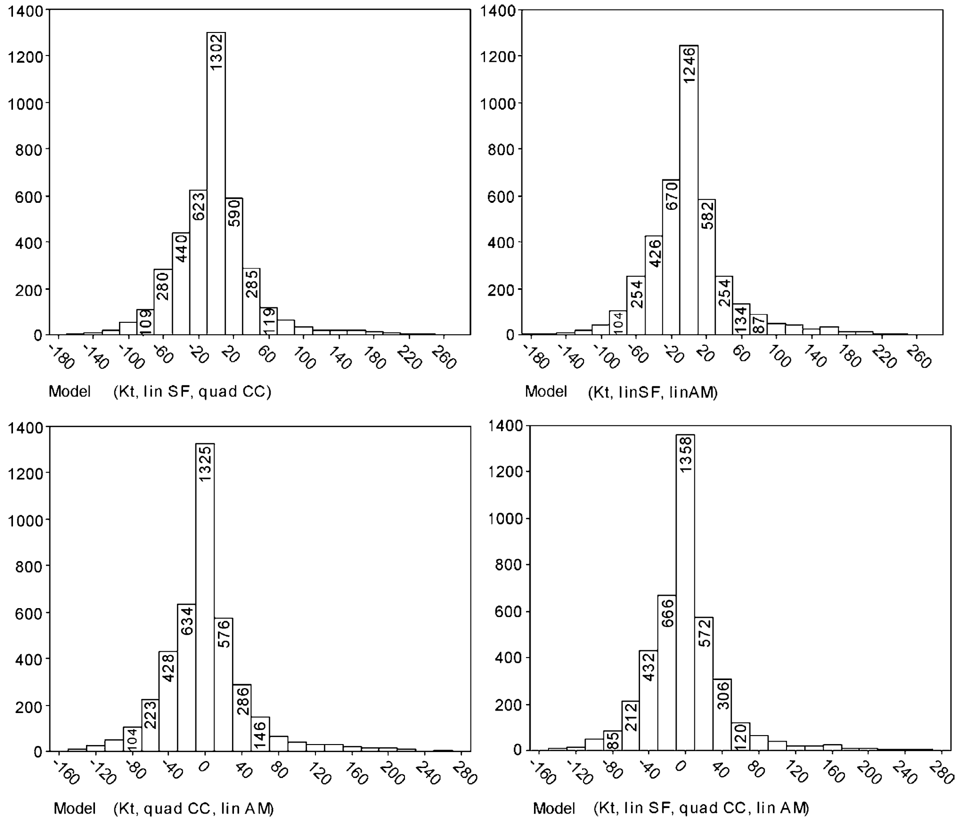


Fig. 23. (*Continued*)

For Spain:

$$k = \left(0.863 + 0.343\text{SF} + 0.037m - 0.004m^2\right) + \left(1.005 - 2.781\text{SF} - 0.176m + 0.007m^2\right)k_\text{t} + \left(-2.259 + 2.843\text{SF} + 0.016m - 0.020m^2\right)k_\text{t}^2$$

For UK:

$$k = \left(0.899 - 0.683\text{SF} + 0.648\text{SF}^2 + 0.028m - 0.002m^2\right) + \left(0.880 - 0.666\text{SF} - 0.314\text{SF}^2 - 0.158m + 0.003m^2\right)k_\text{t} + \left(-1.751 + 2.786\text{SF} - 1.924\text{SF}^2 + 0.044m + 0.012m^2\right)k_\text{t}^2$$

2. $k = f(k_\text{t}, \text{CC}, m)$—for locations that provide information on global radiation and cloud cover only.

For India:

$$k = (1.068 - 0.016\mathrm{CC} + 0.0001\mathrm{CC}^2 - 0.012m - 0.004m^2) + (-0.897 + 0.297\mathrm{CC} + 0.003\mathrm{CC}^2 - 1.143m + 0.196m^2)k_t + (-0.951 - 0.286\mathrm{CC} - 0.003\mathrm{CC}^2 + 2.109m - 0.319m^2)k_t^2$$

For Spain:

$$k = (1.124 - 0.018CC - 0.006m + 0.001m^2) + (-1.905 + 0.272CC + 0.022m - 0.012m^2)k_t + (0.822 - 0.308CC - 0.094m + 0.028m^2)k_t^2$$

For UK:

$$k = (1.266 - 0.007\mathrm{CC} - 0.005\mathrm{CC}^2 + 0.007m - 0.0004m^2) + (-1.469 - 0.124\mathrm{CC} + 0.051\mathrm{CC}^2 + 0.041m - 0.011m^2)k_t + (0.21 + 0.233\mathrm{CC} - 0.065\mathrm{CC}^2 - 0.263m + 0.042m^2)k_t^2$$

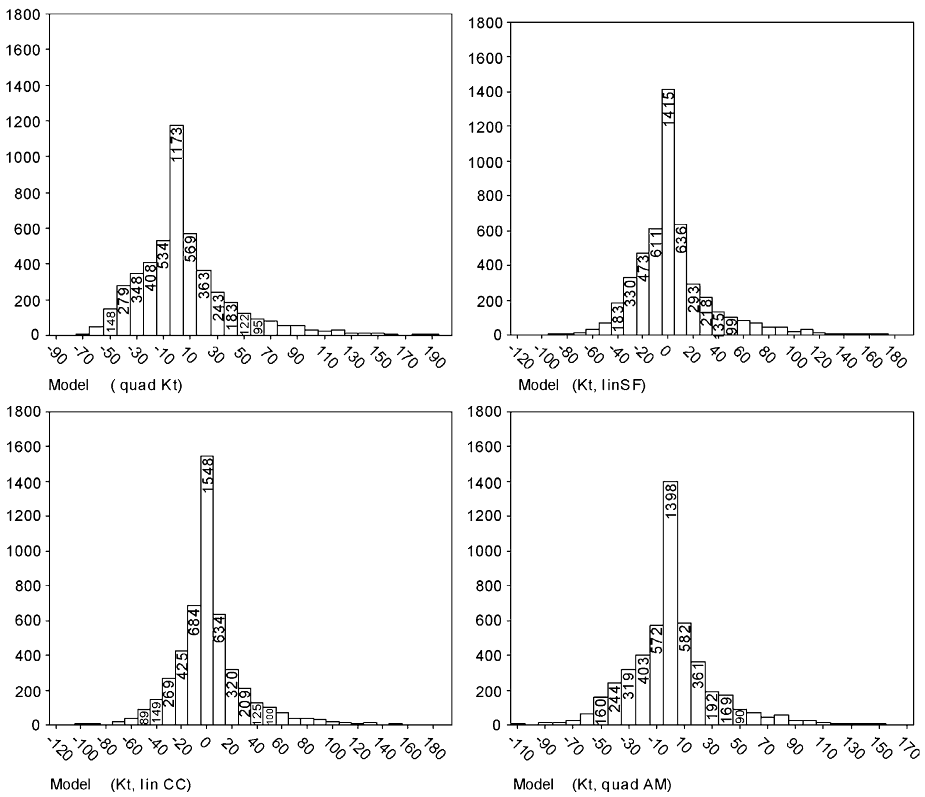


Fig. 24. Error histograms of calculated diffuse radiation of selectively chosen eight models for Gerona.

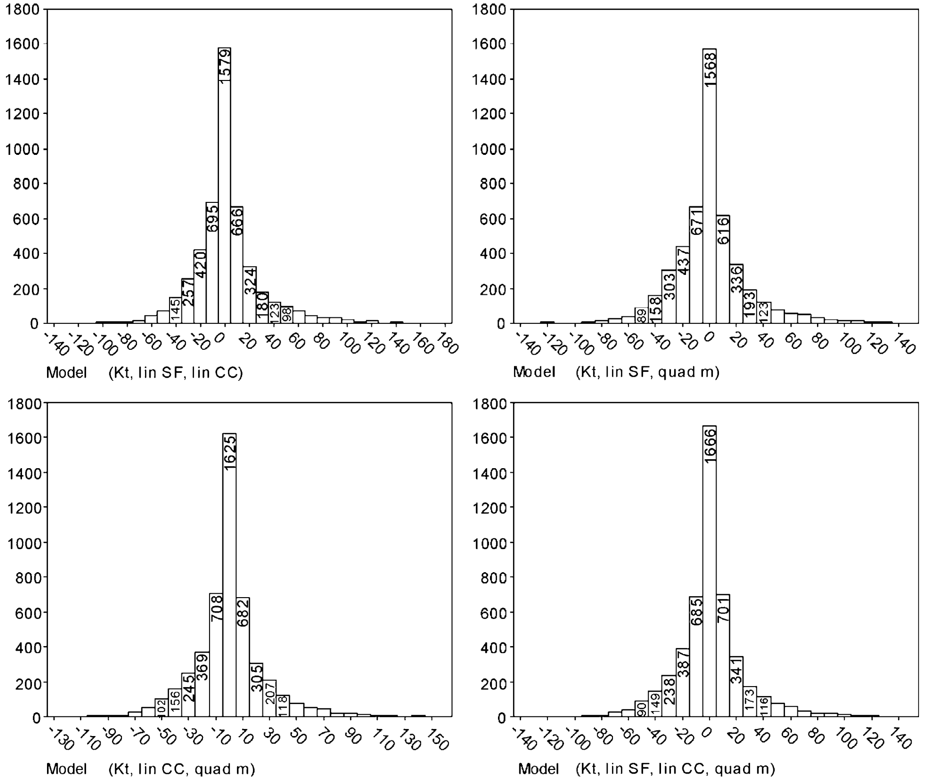


Fig. 24. (*Continued*)

3. $k = f(k_t, \mathrm{SF}, \mathrm{CC}, m)$—for locations that provide information on all the parameters used in this study.

For India:

$$\begin{aligned} k = {} & \left(0.81 - 0.14\mathrm{SF} - 0.009\mathrm{CC} + 0.0001m - 0.006m - 0.004m^2\right) \\ & + \left(0.847 - 1.582\mathrm{SF} + 0.121\mathrm{CC} + 0.001\mathrm{CC}^2 - 0.952m + 0.165m^2\right)k_t \\ & + \left(-2.230 + 1.192\mathrm{SF} - 0.129\mathrm{CC} - 0.001\mathrm{CC}^2 + 1.878m - 0.283m^2\right)k_t^2 \end{aligned}$$

For Spain:

$$\begin{aligned} k = {} & \left(0.981 - 0.132\mathrm{SF} - 0.005\mathrm{CC} + 0.001m + 0.0001m^2\right) \\ & + \left(-0.872 - 1.159\mathrm{SF} + 0.175\mathrm{CC} + 0.024m - 0.009m^2\right)k_t \\ & + \left(-0.226 + 1.213\mathrm{SF} - 0.203\mathrm{CC} - 0.100m + 0.025m^2\right)k_t^2 \end{aligned}$$

Table 10
Accuracy scores of developed models for India

No	Model type	Locations			
		Chennai	Mumbai	New Delhi	Pune
1	$k = f(\text{qkt})$	2.02	2.04	1.38	0.98
2	$k = f(\text{qkt, linSF})$	2.19	2.37	2.74	2.49
3	$k = f(\text{qkt, qSF})$	2.11	2.26	2.78	2.48
4	$k = f(\text{qkt, linCC})$	2.02	2.22	-	3.12
5	$k = f(\text{qkt, qCC})$	2.78	2.36	-	3.09
6	$k = f(\text{qkt, lin m})$	2.17	2.10	1.75	1.17
7	$k = f(\text{qkt, q m})$	2.34	2.57	2.57	1.37
8	$k = f(\text{qkt, linSF, linCC})$	2.96	2.35	-	3.05
9	$k = f(\text{qkt, linSF, qCC})$	2.90	2.24	-	2.82
10	$k = f(\text{qkt, qSF, qCC})$	2.84	2.22	-	2.69
11	$k = f(\text{qkt, linSF, lin m})$	3.55	2.92	2.94	3.02
12	$k = f(\text{qkt, linSF, q m})$	3.97	2.98	3.24	2.91
13	$k = f(\text{qkt, qSF, q m})$	3.84	2.97	3.40	3.04
14	$k = f(\text{qkt, qCC, lin m})$	3.34	2.94	-	3.35
15	$k = f(\text{qkt, qCC, q m})$	3.58	3.04	-	3.18
16	$k = f(\text{qkt, linSF, qCC, lin m})$	3.47	3.20	-	3.42
17	$k = f(\text{qkt, linSF, qCC, q m})$	3.79	3.18	-	3.35
18	$k = f(\text{qkt, qSF, qCC, q m})$	3.73	3.20	-	3.41

Table 11
Accuracy scores of developed models for Japan

No	Model type	Locations	
		Fukuoka	Sapporo
1	$k = f(\text{quadkt})$	1.96	2.87
2	$k = f(\text{qkt, linSF})$	2.80	2.44
3	$k = f(\text{qkt, qSF})$	2.89	2.61
4	$k = f(\text{qkt, lin m})$	2.98	2.89
5	$k = f(\text{qkt, q m})$	3.20	2.90
6	$k = f(\text{qkt, linSF, lin m})$	3.20	3.15
7	$k = f(\text{qkt, linSF, q m})$	3.29	3.15
8	$k = f(\text{qkt, qSF, lin m})$	3.23	3.30
9	$k = f(\text{qkt, qSF, q m})$	3.35	3.31

For UK:

$$
\begin{aligned}
k = {} & \left(0.917 - 0.723\mathrm{SF} - 0.741\mathrm{SF}^2 + 0.008\mathrm{CC} + 0.001\mathrm{CC}^2 + 0.023m - 0.001m^2\right) \\
& + \left(-0.604 - 0.012\mathrm{SF} + 1.041\mathrm{SF}^2 + 0.073\mathrm{CC} + 0.012\mathrm{CC}^2 - 0.124m + 0.001m^2\right)k_{\mathrm{t}} \\
& + \left(-1.573 + 1.542\mathrm{SF} - 0.621\mathrm{SF}^2 + 0.105\mathrm{CC} - 0.013\mathrm{CC}^2 + 0.025m + 0.014m^2\right)k_{\mathrm{t}}^2
\end{aligned}
$$

Table 12
Accuracy scores of developed models for Spain

No	Model type	Locations Gerona	Madrid
1	$k = f(\text{qkt})$	1.63	1.40
2	$k = f(\text{qkt}, \text{linSF})$	2.38	2.88
3	$k = f(\text{qkt}, \text{qSF})$	2.43	3.04
4	$k = f(\text{qkt}, \text{linCC})$	2.97	3.78
5	$k = f(\text{qkt}, \text{qCC})$	3.01	3.82
6	$k = f(\text{qkt}, \text{lin m})$	2.87	2.60
7	$k = f(\text{qkt}, \text{q m})$	3.10	3.30
8	$k = f(\text{qkt}, \text{linSF}, \text{linCC})$	3.09	3.85
9	$k = f(\text{qkt}, \text{qSF}, \text{qCC})$	3.12	3.96
10	$k = f(\text{qkt}, \text{linSF}, \text{lin m})$	3.22	3.16
11	$k = f(\text{qkt}, \text{linSF}, \text{q m})$	3.52	3.53
12	$k = f(\text{qkt}, \text{qSF}, \text{q m})$	3.47	3.60
13	$k = f(\text{qkt}, \text{linCC}, \text{lin m})$	3.47	3.75
14	$k = f(\text{qkt}, \text{linCC}, \text{q m})$	3.66	4.02
15	$k = f(\text{qkt}, \text{qCC}, \text{q m})$	3.70	4.06
16	$k = f(\text{qkt}, \text{linSF}, \text{linCC}, \text{lin m})$	3.60	3.74
17	$k = f(\text{qkt}, \text{linSF}, \text{linCC}, \text{q m})$	3.85	3.98
18	$k = f(\text{qkt}, \text{qSF}, \text{qCC}, \text{q m})$	3.85	4.06

Table 13
Accuracy scores of developed models for UK

No	Model type	Location (Bracknell)
1	$k = f(\text{qkt})$	1.79
2	$k = f(\text{qkt}, \text{linSF})$	2.36
3	$k = f(\text{qkt}, \text{qSF})$	2.50
4	$k = f(\text{qkt}, \text{linCC})$	2.96
5	$k = f(\text{qkt}, \text{qCC})$	2.92
6	$k = f(\text{qkt}, \text{lin m})$	3.08
7	$k = f(\text{qkt}, \text{q m})$	3.50
8	$k = f(\text{qkt}, \text{linSF}, \text{linCC})$	3.00
9	$k = f(\text{qkt}, \text{qSF}, \text{qCC})$	3.04
10	$k = f(\text{qkt}, \text{linSF}, \text{lin m})$	3.49
11	$k = f(\text{qkt}, \text{qSF}, \text{q m})$	3.82
12	$k = f(\text{qkt}, \text{qCC}, \text{lin m})$	3.27
13	$k = f(\text{qkt}, \text{qCC}, \text{q m})$	3.62
14	$k = f(\text{qkt}, \text{linSF}, \text{qCC}, \text{lin m})$	3.61
15	$k = f(\text{qkt}, \text{linSF}, \text{qCC}, \text{q m})$	3.87
16	$k = f(\text{qkt}, \text{qSF}, \text{qCC}, \text{q m})$	3.92

8. Conclusions

The all-sky meteorological radiation model was improved with regards to the estimation of the beam component of the horizontal irradiance, which in turn led to improvement in the estimation of broadband global horizontal irradiance. The improvement is judged

based on the reduction of scatter of data points within a computed-measured plot as well as statistical error analysis.

The model was evaluated for the two scenarios of the available data. One scenario is high quality datasets for sites, where hourly synoptic data may be available to the user and the other being low quality datasets that contain only daily records of sunshine and temperature. Using high quality datasets, the estimation was precise and accurate with a very low rate of residuals. However, the low quality datasets were unfortunately less precise and yet the estimation was found to be between 70 and 80% accurate. Users may find the proposed procedure to estimate radiation by using the easily available synoptic information quite useful as the latter datasets are now available on the internet. Using only monthly averaged maximum and minimum dry bulb temperatures, humidity and daily recorded sunshine data, it was possible to estimate the global irradiance with a $\pm$20–30% error.

During the study on the different cloud based radiation models, two existing bi-variate cloud-based solar irradiance models and their two variations have been compared to three newly proposed models. The two proposed models are bi-variates with the second being a hybrid of the first. All the bi-variate models presented in this paper rely on the cloud cover and the solar altitude angle as variables for the estimation of the horizontal solar irradiance components. The third proposed model is a tri-variate model. In addition to cloud cover and the solar altitude angle, this model utilizes air mass as a third variable. The existing models estimate I_G and I_D and do not estimate I_B, in these cases I_B was deduced from the other two components. The proposed models estimate I_G and I_B then, I_D is deduced from global and beam components.

Based on visual observations of the measured vs. calculated I_G, I_D and I_B scatter plots, the estimation error histograms and also using statistical indicators such as slope and R^2 of best fit line between measured and calculated values, MBE and RMSE, kurtosis and skewness of the estimation error histograms, the models were ranked in the following order based on their overall performance using an accuracy scoring system. The older models were found to be less accurate than their respective modified variations based on local conditions. It was found that the Kasten and Czeplak and its variation, the Muneer et al. models are more accurate than the Lam and Li model.

The newly proposed models outperform the existing models in the estimation of I_B and I_D. The improvement of the proposed bi-variate hybrid model and the proposed tri-variate model compared to the proposed bi-variate model cannot be justified considering the extra coefficients, and variables involved in the procedure. Therefore, the simple bi-variate model, M3, proposed in this article, is recommended to be used for the estimation of the solar horizontal irradiance components.

The work presented under diffuse validation modeling demonstrated the influence of synoptic elements, sunshine fraction and cloud cover, and air mass on the regressions used for estimation of hourly diffuse radiation. It was concluded that they have a significant bearing on the model accuracy and can be effectively used along with global radiation to improve the diffuse radiation modeling. The effect of each synoptic parameter was explained qualitatively through a graphical analysis. It was found that both SF and CC have a strong bearing on diffuse ratio. A similar graphical analysis for air mass revealed that although it would not improve the modeling if used alone with k_t, unlike sunshine fraction or cloud cover in some cases, yet it might also lead to better estimation of diffuse irradiance if used in conjunction with any or both of the two parameters. Generally, the

composite models i.e. the models including all the parameters, showed significant improvement in the estimation of diffuse radiation. They were recommended as optimum providing reasonable accuracy without making the model overtly complex. Although, the models were found to be region-specific, it was also noted that model types are fairly consistent for neighboring stations. Country-wise models were proposed to estimate diffuse radiation for all the sites falling within that region with reasonable accuracy. Three such coherent models were developed under each of the above four categories to suit the availability of parameters. It was shown presently that incorporation of the three parameters (SF, CC and m) in addition to the k_t variable yields a significant improvement over the conventional k–k_t models.

References

[1] Hubbard KG. Spatial variability of daily weather variables in high plains of the USA. Agric Meteorol 1994;68:29–41.

[2] Gueymard C. Prediction and performance assessment of mean hourly global radiation. Solar Energy 1999;68:285–301.

[3] Stoffel T, Redo I, Myers D, Renne D, Wilcox S, Treadwell J. Current issues in terrestrial solar radiation instrumentation for energy, climate and space applications. Metroloogia 2000;37(5):399–402.

[4] Muneer T, Younes S, Claywell R. Quality control of solar radiation data: present status and proposed new approaches. Energy 2005;30:1533–49.

[5] Page J. Proposed quality control procedures for the Meteorological Office data tapes relating to global solar radiation, diffuse solar radiation, sunshine and cloud in the UK. Report submitted to CIBSE Guide Solar Data Task Group, Chartered Institute of Building Services Engineers, 222 Balham High Road, London, UK; 1997.

[6] Muneer T, Fairooz F. Quality control of solar radiation and sunshine measurements—lessons learnt from processing worldwide databases. Building Services Eng Res Technol 2002;23(3):151–66.

[7] Muneer T, Gul M, Kambezidis H. Evaluation of an all-sky meteorological radiation model based against long-term measured hourly data. Energy Convers Manage 1998;39:303–17.

[8] Gul M, Muneer T, Kambezidis H. Models for obtaining solar radiation from other meteorological data. Solar Energy 1998;68:99–108.

[9] Muneer T, Gul M. Evaluation of sunshine and cloud cover based models for generating solar radiation data. Energy Convers Manage 2000;41:461–82.

[10] Muneer T. Solar radiation and daylight models. Oxford: Elsevier; 2004.

[11] Kasten F, Czeplak G. Solar and terrestrial radiation dependant on the amount of type of cloud. Solar Energy 1980;24:177–89.

[12] Gul M, Muneer T, Kambezidis H. Models for obtaining solar radiation from other meteorological data. Solar Energy 1998;64:99–108.

[13] Muneer T, Gul M. Evaluation of sunshine and cloud cover based models for generating solar radiation data. Energy Convers Manage 2000;41:461–82.

[14] Lam J, Li D. Correlation analysis of solar radiation and cloud cover. Int J Ambient Energy 1998;19:187.

[15] Kasten F. Discussion on the relative air mass. Light Res Technol 1993;25:129.

[16] Barker H. Solar radiative transfer through clouds possessing isotropic variable extinction coefficient. QJR Meteorol Soc 1992;118:1145–62.

[17] Harrison, Coombes C. Empirical relationship of cloud shade to point cloudiness (Canada). Solar Energy 1986;37:417–21.

[18] Myers D. Solar radiation modeling and measurements for renewable energy applications: data and model quality. Energy 2005;30.

[19] Brinsfield R, Yaramanoglu M, Wheaton F. Ground level solar radiation prediction model including cloud cover effects. Solar Energy 1984;33:493–9.

[20] Bennett. Correlation of daily insolation with daily total sky cover, opaque sky cover and percentage of possible sunshine. Solar Energy 1969;12:391–3.

[21] Coulson K. Solar and terrestrial radiation, methods and measurements. New York: Academic Press; 1975 pp. 60–100.
[22] De Miguel A, Bilbao J, Aguiar R, Kambezidis H, Negro E. Diffuse solar radiation model evaluation in the north Mediterranean belt area. Solar Energy 2001;70(2):143–53.
[23] Gopinathan KK, Soler A. The determination of monthly mean hourly diffuse radiation on horizontal surfaces using equations based on hourly clearness index, sunshine fraction and solar elevation. Int J Solar Energy 1996;18:115–24.